

BTO 2014.046 | Juni 2014

BTO rapport

Strategieën voor optimale plaatskeuze van waterkwaliteits-sensoren in het distributienet

KWR Watercycle Research Institute

BTO Rapport

Strategieën voor optimale plaatskeuze van waterkwaliteitssensoren in het distributienet

BTO 2014.046 | Juni 2014

Opdrachtnummer
400554/022/002

Projectmanager
Stefan Kools

Opdrachtgever
BTO

Kwaliteitsborger(s)
Mirjam Blokker

Auteur(s)
Peter van Thienen

Verzonden aan
Dit rapport is verspreid onder BTO-participanten en is openbaar.

Jaar van publicatie
2014

Meer informatie

T 030 6069 602
E peter.vanthienen@kwrwater.nl

PO Box 1072
3430 BB Nieuwegein
The Netherlands

T +31 (0)30 60 69 511
F +31 (0)30 60 61 165
E info@kwrwater.nl
I www.kwrwater.nl

BTO 2014.046 | Juni 2014 © KWR

Alle rechten voorbehouden.
Niets uit deze uitgave mag worden verveelvoudigd, opgeslagen in een geautomatiseerd gegevensbestand, of openbaar gemaakt, in enige vorm of op enige wijze, hetzij elektronisch, mechanisch, door fotokopieën, opnamen, of enig andere manier, zonder voorafgaande schriftelijke toestemming van de uitgever.

Samenvatting

Gebruik van een modelmatige optimalisatie bij het inrichten van sensornetwerken voor waterkwaliteit levert– bij een gelijk aantal sensoren - beter functionerende sensornetwerken op dan uniforme verspreiding of plaatsing op basis van gevoel en ervaring. Dit kan de pakkans van een besmetting substantieel verhogen en de gemiddelde tijd tot eerste detectie flink verkorten. Vóór het inrichten van een sensornetwerk is het belangrijk de doelstellingen helder te hebben en daarop te optimaliseren. Bovendien is het belangrijk ook met de buitenwacht goed over de te behalen resultaten te communiceren om geen verkeerde verwachtingen te scheppen.

Belang: onderbouwde keuzes en realistische verwachtingen
Monitoring van de waterkwaliteit met sensoren op strategische punten in het leidingnet geeft online zicht op de waterkwaliteit tijdens distributie en nabij de tap. Nu is er meestal pas zicht op de waterkwaliteit enkele dagen na levering en/of op slechts enkele locaties. De prestaties van een waterkwaliteitssensornetwerk zijn sterk afhankelijk van de ruimtelijke configuratie. Omdat het drinkwaternetwerk zeer uitgebreid is en sensoren vaak duur zijn, is het van belang allereerst een duidelijke visie te hebben wat met het netwerk wordt beoogd. Vervolgens is het nodig een onderbouwde strategie te ontwikkelen voor de bepaling van het benodigde aantal sensoren en de optimale locaties voor hun plaatsing. Deze aanpak draagt bij aan een optimaal functionerend sensornetwerk, waarover binnen en buiten het waterbedrijf realistische verwachtingen bestaan ten aanzien van de toegevoegde waarde en de beperkingen van waterkwaliteitssensoren.

Aanpak: beschouwing van doelstellingen, numerieke evaluatie en optimalisatie
De beschreven problematiek is vanuit modelmatige hoek aangevlogen. Eerst is bekeken hoe de prestatie van een sensornetwerk kan worden beschreven en met welke doelstellingen in het achterhoofd een sensornetwerk kan worden ontworpen. Vervolgens is op basis van een combinatie van hydraulische modellen, besmettingsscenario's en een optimalisatiemethodiek (genetisch algoritme) een programma geschreven voor de evaluatie en optimalisatie van ontwerpen van sensornetwerken. Dit programma is eerst getest op het deelgebied Leeuwarden-Bergum van de Proeftuin Noordbergum van Vitens en daar vervolgens ook toegepast. Hierbij zijn optimale configuraties voor diverse doelstellingen bepaald en zijn bovendien onderlinge relaties en trends vastgesteld. De resultaten zijn verwerkt in een beschouwing met aanbevelingen voor het opzetten van een waterkwaliteitssensornetwerk.

Resultaten: betere prestatie door numerieke optimalisatie
De resultaten van de modelberekeningen zijn eenduidig: numerieke optimalisatie zorgt voor een duidelijk beter presterende configuratie van een bepaald aantal waterkwaliteits-sensoren dan wanneer een gelijk aantal sensoren evenredig wordt verspreid of de plaatsing op basis van gevoel en ervaring wordt gekozen (zie Tabel S1). In dit onderzoek is de gemiddelde pakkans bijvoorbeeld bijna verdubbeld en de gemiddelde tijd tot eerste detectie met enkele uren verminderd. De geoptimaliseerde netwerken functioneren in de regel zelfs beter dan een bestaande of uniform verdeelde configuratie op doelstellingen waarvoor ze niet zijn geoptimaliseerd. Hiermee is het nut van een dergelijke numerieke beschouwing overduidelijk aangetoond. Dit wil echter niet zeggen dat een geoptimaliseerd sensornetwerk onder alle omstandigheden goed presteert. Een sensornetwerk dat voor een bepaald doel

wordt ingericht, kan namelijk verre van optimaal functioneren wanneer het ook voor een andere doelstelling wordt ingezet. Pakkans en detectietijd (en in mindere mate gemak van bronbepaling) zijn doelstellingen die in het onderzochte netwerk goed samen opgaan; een goede dekkingsgraad en een goede detectietijd lijken moeilijk samen te gaan voor een beperkt aantal sensoren (zie Tabel S1).

Tabel S1: Prestatie van geoptimaliseerde, huidig bestaande en uniforme sensornetwerken op verschillende doelstellingen in het deelgebied Leeuwarden-Bergum voor 10 (links) respectievelijk 15 (rechts) sensoren. De huidige configuratie (onderste rij, linkervakjes) betreft de opstelling van 10 Eventlab-stations in het gebied. Pakkans en dekkingsgraad zijn weergegeven als fractie van het totaal, de detectietijd is in uren en bronbepaling geeft de gemiddelde omvang van het potentiële brongebied in kilometers weer.

(geoptimaliseerde) configuratie	*prestatie met betrekking tot doelstelling (10/15 sensoren)*									
	pakkans		*detectietijd (uren)*		*dekking*		*redundante dekking*		bron-bepaling	
pakkans	0,41	0,50								
detectietijd			10,4	9,13						
dekking					0,60	0,62				
redundante dekking							0,46	0,48		
bronbepaling									244	238
huidig/uniform	0,22	0,26	13,9	12,5	0,36	0,44	0,08	0,25	292	285

prestatie: slechtst — best

Implementatie: basis voor een bruikbaar sensornetwerk en realistische verwachtingen

Het is van groot belang dat een drinkwaterbedrijf goed nadenkt over de doelstelling van een eventueel te installeren sensornetwerk voor waterkwaliteit, voordat de omvang en configuratie van het netwerk wordt bepaald. Met een beperkt aantal sensoren is het onmogelijk om een netwerk volledig te beschouwen en alle risico's af te dekken. Dit moet zeker ook worden meegewogen in de communicatie met de buitenwereld. Dat neemt niet weg dat waterkwaliteitssensoren hoe dan ook informatie opleveren die *kan* leiden tot een snellere constatering van een besmetting en een meer doelgerichte aanpak van besmettingen: zowel bij het beperken van verdere verspreiding en het verwijderen van de besmetting als bij de communicatie richting klanten. De gebruikte en ontwikkelde methodieken voor optimale locatiekeuze van waterkwaliteitssensoren maken het mogelijk om verschillende doelstellingen tegen elkaar af te wegen en tot een realistische verwachting van de opbrengsten van een dergelijk netwerk te komen. Dit draagt bij aan zowel de interne besluitvorming als de externe communicatie van beslissingen en implementaties op dit gebied. Zo kan een afweging worden gemaakt tussen kosten en geboden mate van bescherming en leveren de uitkomsten van deze berekeningen randvoorwaarden voor de respons die een bedrijf kan nastreven.

Inhoud

	Samenvatting	3
	Inhoud	5
1	Inleiding	7
1.1	Monitoring van waterkwaliteit in het distributienet met sensoren	7
1.2	Succescriteria voor sensornetwerken voor waterkwaliteitsmonitoring	8
1.3	Huidige stand van zaken van optimale locatiekeuze voor sensoren	11
1.4	Leeswijzer	11
2	Methodiek: optimalisatie, **doelstellingen en besmettingsscenario's**	12
2.1	Inleiding	12
2.2	Benadering	13
2.3	Optimalisatiemethodiek	13
2.4	Instellingen en keuzes voor het genetische algoritme	15
2.5	Doelstellingen van optimalisatie	15
2.6	Uitgangspunten bij optimalisatiedoelstellingen	17
2.7	Besmettingsscenario's	19
2.8	Potentiële sensorlocaties	20
2.9	Representatie van waterkwaliteitssensoren en –bepalingen	20
2.10	Aannames	20
2.11	Gebruikte software	20
2.12	Ideeën voor verdere ontwikkeling	21
3	Resultaten	22
3.1	Inleiding	22
3.2	Proefgebied en hydraulische modellen	22
3.3	Potentiële besmettings- en sensorlocaties	23
3.4	Criterium 1a: Dekkingsgraad	24
3.5	Criterium 1b: Redundantie in de dekkingsgraad	25
3.6	Criterium 2: Pakkans	26
3.7	Criterium 3: Detectietijd	27
3.8	Criterium 4: Gemak van bronbepaling	28
3.9	Criterium 5a: Robuustheid m.b.t. falen van sensoren: kunnen detecteren (pakkans)	28
3.10	Vergelijking van criteria	30
3.11	Effect van selectie potentiële locaties	32
4	Beschouwing en implementatie	35
4.1	Visie en doelstellingen	35
4.2	Rekenwerk en mensenwerk	35
4.3	Verwachtingen met betrekking tot sensoring in distributienetwerken	36
4.4	Hoe verder te gaan met sensoring van waterkwaliteit in distributienetwerk	36
4.5	Naar een business case voor sensoring van waterkwaliteit	37
4.6	Aanbevelingen en randvoorwaarden	38
5	Conclusies en aanbevelingen	39
5.1	Conclusies	39

5.2 Aanbevelingen voor implementatie 40
5.3 Aanbevelingen voor vervolgonderzoek en verdere ontwikkeling 41

6 Verklarende woordenlijst 42

7 Literatuur 43

Bijlage I: Algoritmen 45
Bijlage II: Instellingen en evaluatie van optimalisatiedoelstellingen en -algoritmen 50
- Instelling van optimalisatie-algoritme 59

Bijlage III: Beschrijving en benchmark van ontwikkelde tool voor optimale sensorlocatie-keuze 65
- **Beschrijving van voorbereidende en ondersteunende programma's** 65
- Beschrijving van de Contamination Source Toolkit 65
- Stappenplan voor evaluatie en optimalisatie met cst en bijbehorende tools 69
- Benchmark 69

1 Inleiding

1.1 Monitoring van waterkwaliteit in het distributienet met sensoren

Monitoring van de waterkwaliteit op strategische punten in het leidingnet met behulp van sensoren geeft online zicht op de waterkwaliteit tijdens distributie en nabij de tap, en niet, zoals nu het geval is, pas enkele dagen na levering en/of op slechts enkele locaties. Omdat het drinkwaternetwerk zeer uitgebreid is en sensoren vaak duur zijn, is het van belang een onderbouwde strategie te ontwikkelen voor de bepaling van het benodigde aantal sensoren en de optimale locaties voor plaatsing hiervan.

Er bestaat een verzameling van wetenschappelijke literatuur over de optimale keuze van monster- en sensorlocaties (o.a. Bush en Uber, 1998, Speight et al., 2004, Kapelan et al., 2005) Het aantal implementaties van de hierin beschreven methodieken in voor waterbedrijven toepasbare programma's is echter beperkt. Bovendien richten deze zich vaak specifiek op het minimaliseren van de effecten van een eventuele chemische of microbiologische besmetting (bijvoorbeeld TEVA-SPOT: Threat Ensemble Vulnerability Assessment – Sensor Placement Optimization Tool (Berry et al., 2010). Zij optimaliseren de plaatskeuze van sensoren specifiek om dit doel te bewerkstelligen. De respons van een bedrijf op de constatering van een besmetting bepaalt mede hoe groot de effecten uiteindelijk zijn. Deze respons wordt op een eenvoudige wijze in dergelijke optimalisaties meegenomen, maar blijkt bijna allesbepalend voor de resultaten te zijn.

Sensoren kunnen echter ook met andere doelstellingen dan het detecteren van eventuele besmettingen in het net worden geplaatst, zoals bijvoorbeeld ter controle/beheersing van de bedrijfsvoering of om "te weten wat er gebeurt in het leidingnet". Bovendien werken niet alle sensoren op een voor incidentdetectie relevante tijdschaal (bijvoorbeeld de Continuous Biofilm Monitor, maar ook monsternames met waterkwaliteitsanalyse in het lab kunnen in deze context als discontinu opererende sensor met een lange responstijd worden beschouwd). Ook kan in het ene geval de absolute waarde van een parameter van belang zijn, terwijl in het andere geval juist een verandering van de parameter veelzeggend is, maar het absolute niveau van de parameter niet. Deze verschillende doelstellingen, toepassingen en sensoraarden en de genoemde visie op het belang van de respons vergen een bredere benadering van de optimalisatie.

Dit project beoogt een geschikte optimale plaatsingsmethodiek te ontsluiten voor brede toepassing door de Nederlandse drinkwaterbedrijven. Hierbij worden optimalisatie-doelstellingen bewust "dicht bij het netwerk" gedefinieerd, en wordt geprobeerd om zachte maar allesbepalende parameters te vermijden. Het begrip sensor wordt hier breed beschouwd, zodat bijvoorbeeld ook de eerder genoemde "langzame" sensoren en bemonstering gevolgd door labanalyses (optimale monsterlocaties) eronder vallen. Hiertoe zijn eerst geschikte algoritmen en benaderingen geselecteerd uit de literatuur, waar nodig verder ontwikkeld en vervolgens geïmplementeerd in een onderzoekstool.

Vanwege voorkomende problemen met de robuustheid van sensoren, onderbrekingen in communicatie en/of opslag, etc. is oog voor datakwaliteit en correcte omgang met de mogelijkheid van fout-positieven en fout-negatieven van groot belang. Met name fout-negatieven zijn uit den bozen, omdat stoffen ongemerkt door het netwerk worden verspreid en door klanten worden geconsumeerd. In dit project wordt rekening gehouden met de

feilbaarheid van sensornetwerken, en wordt het effect van en de gevoeligheid voor fout-negatieve bepalingen expliciet beschouwd, zodat de robuustheid van een sensornetwerk als geheel wordt gewaarborgd.

Met de hierboven genoemde methodieken kan een sensornetwerk worden ontworpen waarmee een bepaald, vooraf gesteld doel kan worden gerealiseerd. Een voorbeeld is een gekwantificeerde vermindering van het risico op het drinken van besmet water door de klanten. Zo kunnen realistische verwachtingen van sensornetwerken worden verkregen, en kan ook de meerwaarde van extra sensoren in een sensornetwerk worden afgewogen tegen de bijkomende kosten. Bovendien wordt duidelijk aan welke randvoorwaarden het distributienetwerk en de kennis van het waterbedrijf daarvan moeten voldoen, om dit doel te kunnen bereiken.

Veel Nederlandse drinkwaterbedrijven hebben nog geen duidelijke strategie of visie met betrekking tot waterkwaliteitssensoren in het distributienet. Voor deze bedrijven kunnen numerieke studies een basis bieden voor realistische verwachtingen met betrekking tot sensornetwerken. Hiermee kan de vorming van een visie en beleid m.b.t. deze sensoren worden ondersteund.

De in dit rapport uitgewerkte ideeën zijn gevoed vanuit de drinkwaterbedrijven via de projectgroep. Deze werd, naast de auteur, gevormd door Bendert de Graaf (Vitens), Maurice van de Roer (Dunea), Peter Schaap (PWN) en Veerle Sperber (Brabant Water).

1.2 Succescriteria voor sensornetwerken voor waterkwaliteitsmonitoring

Kroll en King (2010) beschouwen sensorsystemen voor de detectie van anomalieën in de waterkwaliteit in het distributienetwerk (EDS: event detection systems) in brede zin. Zij doen dit vanuit een militaire achtergrond, maar met de drinkwatervoorziening als toepassing in gedachten. Zij hebben hierbij een lijst van succescriteria voor EDS opgesteld:

(KK1)	Toegevoegde waarde op twee fronten	Naast de detectie van (naar men mag veronderstellen zeldzame) serieuze anomalieën moet het systeem ook voor dagelijks gebruik nuttige informatie opleveren, zoals bijvoorbeeld t.b.v. procesoptimalisatie of het opbouwen/versterken van het consumentenvertrouwen. Hiermee is het waarschijnlijker dat de investering wordt terugverdiend.
(KK2)	Vereisten voor gewenste detectieklasse	Er kunnen drie detectieklassen worden gedefinieerd, namelijk *detectie ter behandeling*, *detectie ter bescherming* en *detectie ter waarschuwing*. De eerste klasse omvat systemen met een zeer hoge betrouwbaarheid, waarop direct behandeling van blootgestelde personen kan worden gebaseerd. De tweede klasse omvat systemen die nog steeds een hoge betrouwbaarheid hebben, zodat zonder verdere bevestiging maatregelen kunnen worden genomen om verdere blootstelling van de bevolking te beperken. De derde klasse omvat minder betrouwbare systemen, die worden gebruikt om blootstelling van de bevolking te beperken terwijl aanvullende bevestigende tests worden uitgevoerd. Van de derde naar de eerste klasse nemen in de regel de kosten toe en de reactietijd af.

(KK3)	Dekkings-karakteristieken	
(KK3.1)	*Kosten*	Wanneer grote gebieden gemonitord moeten worden, zullen budgettaire beperkingen resulteren in een onvolledige dekking.
(KK3.2)	*Beschermings-gebied*	Lokale kenmerken van het te beschermen gebied (bijv. hydraulica van het netwerk en logistieke kwesties) kunnen het moeilijk of onmogelijk maken volledige bescherming te bieden.
(KK3.3)	*Communicatie*	Maatregelen moeten worden genomen om de betrouwbare overdracht van informatie van de verschillende meetpunten naar een centraal punt te waarborgen, waar mensen vervolgens de interpretatie en vervolgstappen verzorgen.
(KK4)	Operationele karakteristieken	
(KK4.1)	*Gebruiksgemak*	Operationeel personeel met beperkte training moet in crisissituaties gemakkelijk van het systeem gebruik kunnen maken.
(KK4.2)	*Mate van automatisering*	Menselijk ingrijpen is alleen wenselijk ten behoeve van onderhoud en reparatie.
(KK4.3)	*Continuïteit van werking*	Onderbrekingen die lang genoeg duren om een besmetting onopgemerkt door het distributiesysteem te laten gaan, zijn onwenselijk.
(KK4.4)	*Betrouwbaarheid*	
(KK4.5)	*Kosteneffectiviteit*	De kosten van het monitoringsysteem moeten vergelijkbaar of minder zijn dan die van het reguliere, reeds bij het waterbedrijf geïmplementeerde monsterprogramma.
(KK5)	Prestatie-karakteristieken	
(KK5.1)	*Breed spectrum*	Idealiter kan het systeem essentiële categorieën van stoffen detecteren (o.a. verschillende klassen pesticiden en herbiciden, aardolieproducten, zware metalen, etc.)
(KK5.2)	*Snelle reactietijd*	Deze moet worden gezien in relatie tot de snelheid waarmee schadelijke doses bij de klant worden afgeleverd (functie van concentratie, stroomsnelheid en gevaarlijke dosis).
(KK5.3)	*Specificiteit*	Aangezien de hoeveelheid stoffen die men zou willen kunnen detecteren immens is, is het toepassen van stofspecifieke sensoren praktisch onhaalbaar. Met een beperkte verzameling van zogenaamde orthogonale sensoren (voor niet-lineair gecorreleerde of onafhankelijke parameters) kan met behulp van mathematische technieken in principe onderscheid maken tussen (klassen van) stoffen, bij beschikbaarheid van voldoende data.
(KK5.4)	*Reproduceer-baarheid*	Reproduceerbaarheid van metingen kan in het lab worden aangetoond voor een systeem.
(KK5.5)	*Weinig fout-positieven en fout-negatieven*	Fout-negatieve signalen zijn minder waarschijnlijk in systemen die meerdere typen sensoren combineren. Fout-positieve metingen zijn meestal het gevolg van hetzij ruis hetzij schaarste aan meetgegevens, en dus in zekere mate

		onvermijdelijk. Daarom moet worden besloten wat een acceptabele frequentie van optreden van fout-positieve metingen is. Naar men verwacht (en hoopt) zal het aantal fout-positieve metingen groter zijn dan het aantal daadwerkelijke besmettingen.
(KK5.6)	*Kwalitatieve indicatie*	Vanwege het mogelijke ontbreken van kennis over specifieke stoffen bij de mensen die met de informatie uit het systeem aan de slag moeten gaan, moet het systeem ook een meer algemene classificatie geven.
(KK5.7)	*Kwantitatieve indicatie*	Idealiter bepaalt het systeem ook concentraties.
(KK5.8)	*Gevoeligheid*	Het systeem moet gevoelig genoeg zijn om schadelijke hoeveelheden van stoffen te kunnen detecteren. Dit is sterk afhankelijk van het lokale ruisniveau.

Veel van de bovengenoemde punten gaan over de sensor zelf. Enkele betreffen het sensornetwerk. Dit rapport behandelt de optimale locatiekeuze voor waterkwaliteitssensoren, en beschouwt daarmee dat deel van het functioneren van een sensorsysteem voor waterkwaliteitsbewaking. De genoemde succescriteria waarbij de plaatskeuze een rol speelt, zijn de volgende:

(KK1)	Toegevoegde waarde op twee fronten	Zoals in de volgende hoofdstukken zal blijken, kunnen vele optimalisatiedoelstellingen voor de plaatskeuze worden gekozen. Voor elk van deze doelstellingen wordt een configuratie gevonden, waarvoor deze het best presteert *voor de gekozen doelstelling*. Voor andere, eveneens relevante doelstellingen moet dan met een suboptimale prestatie genoegen worden genomen. Hierin is echter wel een compromis te vinden. Dit is echter niet geheel wat Kroll en King bedoelen met twee fronten: zij benoemen expliciet het kunnen gebruiken van het sensornetwerk voor monitoring (ten behoeve) van de bedrijfsvoering naast incidentdetectie. Dit laatste aspect wordt in dit rapport nog niet uitgewerkt, maar is een interessant en relevant onderwerp voor mogelijk vervolgonderzoek.
(KK3.1)	*Kosten*	Uitgegaan wordt van een beperkt aantal sensoren. De meerwaarde (in termen van bijvoorbeeld extra bescherming) van extra sensoren kan worden gekwantificeerd, zodat een afweging tussen kosten en opbrengsten kan worden gemaakt.
(KK3.2)	*Beschermings-gebied*	De hydraulica zijn hierop van grote invloed, en deze vormen de basis van de berekeningen in dit rapport.
(KK5.5)	*Weinig fout-positieven en fout-negatieven*	Het effect van fout-negatieve metingen op de uiteindelijke bepaling (van bijvoorbeeld de aanwezigheid van een verontreiniging) en de gevoeligheid van een sensornetwerk hiervoor wordt beschouwd. Andersom biedt dit een handvat om kwantitatieve eisen aan sensoren te formuleren wat betreft het optreden van fout-positieven en fout-negatieven.

1.3 Huidige stand van zaken van optimale locatiekeuze voor sensoren

In het eerste decennium van de 21e eeuw zijn vele methodieken ontwikkeld voor de bepaling van optimale sensorlocaties. Deze benaderingen en hun prestaties zijn voor een groot deel beschreven in Van Thienen en Vogelaar (2012). Dat rapport behandelt eveneens de toepassing van het programma TEVA-SPOT op sensorplaatsingsvraagstukken. Beide onderwerpen worden in dit rapport daarom niet opnieuw besproken.

1.4 Leeswijzer

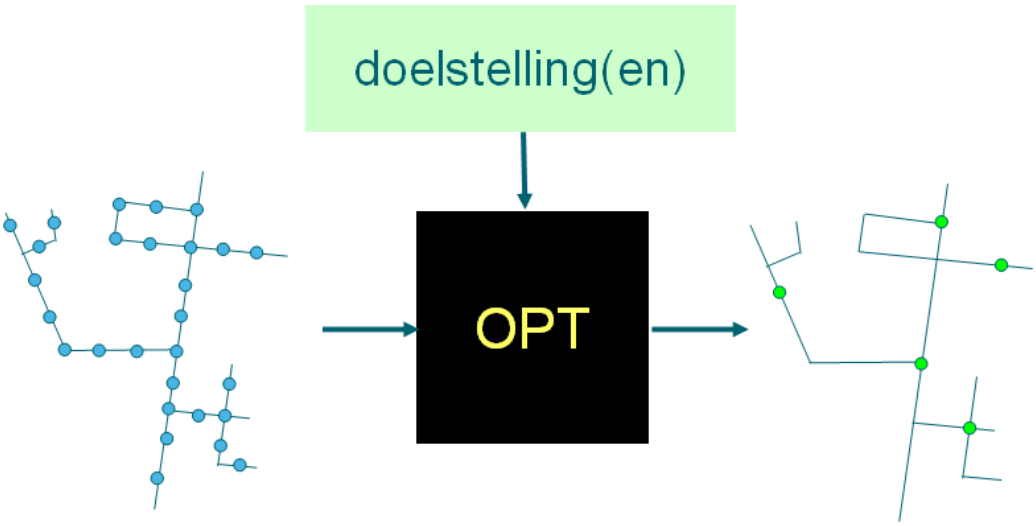
Dit rapport beoogt niet alleen de technisch goed geïnformeerde lezer te bedienen, maar een breder publiek met belangstelling voor toepassing van waterkwaliteitssensoren in drinkwaterdistributienetten te informeren. Daarom zijn technische details zoveel mogelijk in de appendices geplaatst – zij konden vanwege de aard van het onderwerp echter niet altijd vermeden worden in de hoofdtekst. Voor toelichting bij gebruikt jargon is er een verklarende woordenlijst (p. 42).

Hoofdstuk 2 behandelt de gebruikte methodieken, bespreekt uitgebreid de verschillende doelstellingen van optimalisatie en gaat in op besmettingsscenario's, representativiteit en andere aannames. In hoofdstuk 3 worden de resultaten van toepassing van de methodiek op het netwerkmodel van het deelgebied Leeuwarden-Bergum in de Proeftuin Noordbergum gepresenteerd en besproken. Op basis van deze resultaten wordt in hoofdstuk 4 een beschouwing van implementatie van waterkwaliteitssensoren in drinkwaterdistributienetten gegeven. Ten slotte biedt hoofdstuk 5 een overzicht van de belangrijkste bevindingen en aanbevelingen.

2 Methodiek: optimalisatie, doelstellingen en **besmettingsscenario's**

2.1 Inleiding

Optimalisatie is het uit alle mogelijke oplossingen vinden van die oplossing voor een probleem die het best beantwoordt aan de doelstelling van de optimalisatie. Figuur 1 illustreert dit op conceptuele wijze. Stel dat we een netwerk hebben waarin we 5 sensoren gaan plaatsen. Met deze vijf sensoren willen we, wanneer er op een willekeurige plaats in het netwerk een besmetting optreedt, deze besmetting zo snel mogelijk kunnen detecteren (doelstelling). We hebben ongeveer 30 punten in het netwerk waar we deze sensoren om praktische redenen goed zouden kunnen installeren (potentiële sensorlocaties). Een optimalisatie-algoritme selecteert uit de populatie van potentiële locaties die 5 locaties waarmee de gemiddelde detectietijd zo kort mogelijk is (groene stippen).

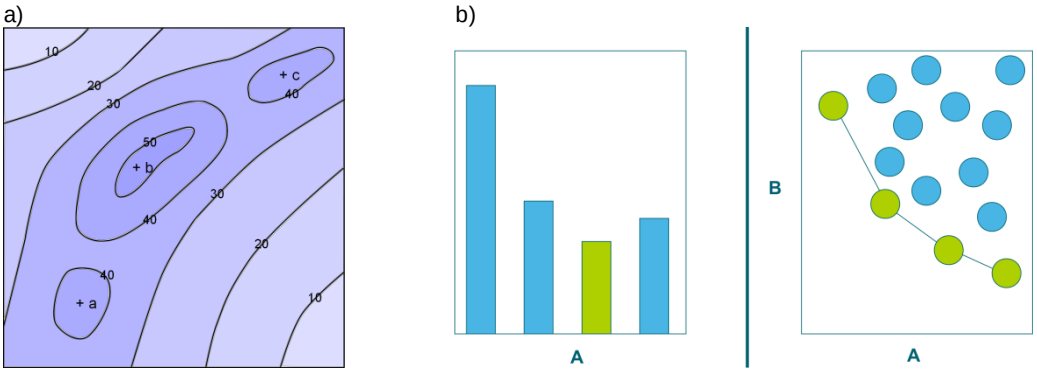


Figuur 1: Aan de hand van een of meer doelstellingen kiest het optimalisatie-algoritme uit een verzameling potentiële locaties die set van locaties die het beste aan de doelstelling beantwoordt. Overgenomen uit Van Thienen en Vogelaar (2012).

Het vinden van de beste oplossing (het globale optimum) kan een uitdaging vormen. Zo worden veel problemen gekenmerkt door meerdere lokale optima naast een enkel globaal optimum. Figuur 2a biedt hiervan een voorbeeld. Dit probleem wordt beschreven door 2 parameters, uitgezet langs de horizontale en verticale as. De prestatie van het systeem als functie van deze twee parameters wordt weergegeven met contouren. We zien bij punten a en c lokale optima, terwijl het globale optimum zich bij punt b bevindt.

Optimalisaties worden, vanwege het relatieve gemak, vaak uitgevoerd op een enkele doelstelling of prestatieparameter. Hierbij kunnen de prestaties van verschillende oplossingen gemakkelijk naast elkaar worden gezet, en de beste geselecteerd (Figuur 2b). Soms is het interessant om te optimaliseren op meerdere doelstellingen – het netwerk moet zo goedkoop mogelijk en tegelijk zoveel mogelijk besmettingen detecteren. Uiteindelijk

moet de mens het relatieve belang van deze twee doelstellingen kiezen. De verzameling van optimale oplossingen voor deze combinatie van doelstellingen wordt weergegeven als een zogenaamd Paretofront in een diagram waarin de prestatie voor de eerste doelstelling op de horizontale as staat weergegeven en de prestatie voor de tweede op de verticale (Figuur 2c, uitbreiding naar meer dimensies is mogelijk).



Figuur 2: a) Illustratie van lokale (punten a en c) en het globale optimum (punt b). De contouren geven de waarde van de prestatieparameter weer. De horizontale en verticale as representeren twee variabelen die de prestatieparameter beïnvloeden. b) Optimalisatie op een enkel criterium A (links) en op twee criteria A en B (rechts). In het eerste geval is verticaal de waarde van de prestatieparameter weergegeven, en wordt de optimale waarde direct uitgelezen (voor een minimalisatie). In het tweede geval zijn optimale (groen) en suboptimale (blauw) combinaties van de parameters A en B weergegeven. De rij van optimale combinaties vormt het zogenaamde Paretofront. Beide deelfiguren zijn overgenomen uit Van Thienen en Vogelaar (2012).

2.2 Benadering

Optimalisatieproblemen lenen zich bij uitstek voor oplossing op numerieke wijze. Het functioneren van het systeem, in dit geval het drinkwaterdistributiesysteem en het daaraan gekoppelde netwerk van waterkwaliteitssensoren, moet hiertoe worden gevat in een numeriek model. Voor het drinkwaterdistributiesysteem voldoen bestaande hydraulische modellen, waaraan voor het modelleren van het sensorsysteem het transport van opgeloste stoffen moet worden toegevoegd (bestaande functionaliteit in veel hydraulische pakketten). Tevens moet de respons van een sensor op lokale concentraties en/of veranderingen worden beschreven. Ten slotte is het in een aantal gevallen ook nodig om de situaties waarin het sensornetwerk moet presteren te beschrijven. De hier summier beschreven benadering wordt gevolgd in dit rapport. De hier genoemde aspecten worden hieronder verder uitgewerkt.

2.3 Optimalisatiemethodiek

Er zijn bijzonder veel wiskundige methoden beschikbaar om een of de optimale configuratie voor een probleem te vinden. Sommige zijn formeel en analytisch, andere zijn heuristisch (d.w.z. op ervaring gebaseerde technieken) of metaheuristisch (op ervaringen gebaseerde technieken zonder veronderstellingen over het op te lossen probleem). Sommige blijven gemakkelijk hangen in lokale optima (d.w.z. kleine variaties van elk van de parameters leiden tot een slechtere oplossing, maar er zijn wel betere oplossingen te vinden wanneer de parameters meer worden gevarieerd, zie Figuur 2a). De ene methodiek is geschikter voor continue problemen, terwijl de andere zich beter leent voor discrete problemen.

De optimalisatieproblemen die we hier willen oplossen, zijn alle discreet: uit een beperkt aantal mogelijkheden moet een deelverzameling van sensoren worden gekozen. De omvang van de op te lossen problemen is in veel gevallen groot. Daarom wordt in dit rapport gebruik

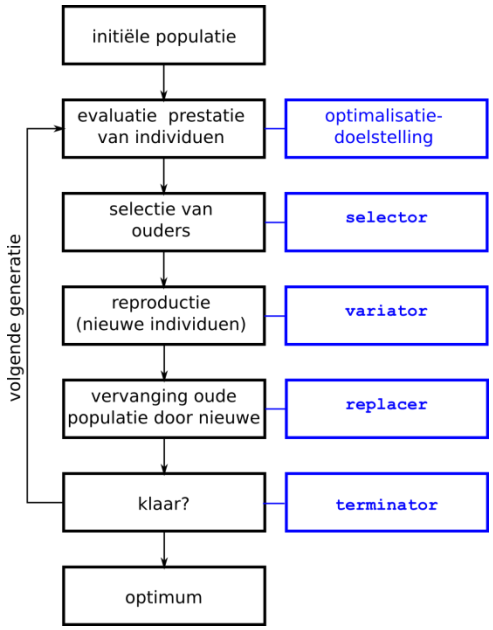
gemaakt van een metaheuristiek die goed kan omgaan met discrete problemen, namelijk een *Genetisch Algoritme.* Een alternatief, in principe eveneens geschikt algoritme is de zogenaamde *Particle Swarm Optimization*. Deze is in dit rapport slechts in beperkte mate gebruikt, maar wordt genoemd omdat deze interessante mogelijkheden biedt voor toekomstige toepassing. Beide technieken, en de gebruikte of bruikbare implementatie, staan beschreven in Tabel 1.

Bij (meta)heuristische technieken is het niet formeel te controleren of een gevonden oplossing werkelijk het (globale) optimum is. Als referentiemethodiek is daarom ook de zogenaamde *Brute Kracht*-benadering geïmplementeerd. Deze rekent simpelweg alle mogelijkheden door en bepaalt dan welke de beste is. Deze methodiek is alleen toepasbaar op kleine problemen, omdat de benodigde rekentijd anders onpraktisch of onhaalbaar wordt.

Tabel 1: Overzicht van gebruikte optimalisatie-algoritmen en implementaties.

algoritme	Genetisch algoritme	Particle Swarm Optimization	Brute Kracht
type	metaheuristisch	metaheuristisch	analytisch
omschrijving	via een proces van selectie en mutatie evolueert een populatie kandidaat-oplossingen richting een optimum, analoog aan biologische evolutie	een populatie van kandidaat-oplossingen beweegt door de zoekruimte volgens simpele regels m.b.t. de positie en snelheid van de kandidaten, die worden beïnvloed door (continu bijgewerkte) lokale en globale op ieder moment bekende beste oplossingen	ieder mogelijke oplossing wordt expliciet door-gerekend; vervolgens wordt de beste geselecteerd
gebruikte implemen-tatie	Genetic Algorithm in *inspyred*	ALPSO in *PyOpt*	eigen implementatie
referenties	http://inspyred.github.io/	http://www.pyopt.org Jansen en Perez (2011)	

De werking van een genetisch algoritme is schematisch weergegeven in Figuur 3. Hierin zijn ook de verschillende componenten weergegeven, die in de volgende paragraaf worden beschreven. De werking is analoog aan (en geïnspireerd op) biologische evolutie. Er wordt begonnen met een populatie van kandidaat-oplossingen (*individuen*) die willekeurig zijn gegenereerd. Dit kunnen bijvoorbeeld 100 configuraties van 10 sensoren met elk een willekeurig toegewezen locatie zijn. Van ieder van deze individuen wordt de prestatie met betrekking tot de optimalisatiedoelstelling berekend. De best presterende individuen laat men zich voortplanten, waarbij de nieuwe individuen varianten van hun ouders zijn. De variatie wordt verkregen door mutatie en combinatie van ouders. Vervolgens wordt een nieuwe populatie van individuen verkregen. Hiertoe worden de best presterende ouders behouden (vastgestelde fractie van het totaal), aangevuld met nieuw verkregen kinderen. Ten slotte wordt bepaald of het proces opnieuw moet worden toegepast op de nieuw verkregen generatie, of dat het proces wordt beëindigd en de beste gevonden oplossing het globale optimum wordt verondersteld te zijn.



Figuur 3: Stappen en componenten van het genetische algoritme. De weergegeven componenten worden beschreven in Tabel 2.

2.4 Instellingen en keuzes voor het genetische algoritme

De mate waarin een genetisch algoritme tot goede oplossingen kan komen, hangt erg af van probleemspecifieke instellingen van het algoritme. Deze paragraaf geeft een summier overzicht van de gemaakte keuzes, met enige uitleg. Een uitgebreidere beschrijving van de instellingen en keuzes, inclusief een numerieke onderbouwing en validatie van deze keuzes kan worden teruggevonden in Bijlage II. Globale keuzes met betrekking tot de verschillende in Figuur 3 genoemde componenten van het genetische algoritme worden beschreven in Tabel 2.

2.5 Doelstellingen van optimalisatie

De uiteindelijke doelstelling van netwerken van waterkwaliteitssensoren is meestal het minimaliseren van de consequenties van een eventuele besmetting van het drinkwater in het leidingnet voor de klanten. In op modellen gebaseerde optimalisaties wordt deze doelstelling vaak uitgedrukt als een van de volgende doelstellingen:

- minimaliseren van tijd tussen besmetting en detectie ervan;
- minimaliseren van het aantal mensen dat wordt blootgesteld;
- minimaliseren van het volume besmet water dat wordt geconsumeerd;
- maximaliseren van de kans op detectie.

Bij berekening van de blootstelling van de bevolking (aantal mensen, geconsumeerd volume) worden aannames gedaan over het gedrag van mensen en over de tijd die nodig is na de eerste detectie om de consumptie stil te leggen. Met name deze laatste factor blijkt vaak bepalend te zijn voor zowel het aantal mensen dat wordt blootgesteld als de optimale sensorconfiguratie die dit aantal moet minimaliseren. Enerzijds is dit goed nieuws, want door zijn reactietijd terug te brengen kan een waterbedrijf effectief de bescherming van zijn klanten vergroten. Anderzijds stelt dit vraagtekens bij de bruikbaarheid van de optimalisatie met deze doelstelling.

Tabel 2: Onderdelen van de GA-optimalisatie en daarvoor in dit rapport gemaakte keuzes

onderdeel		*functie*	*keuze*	*omschrijving/motivering*
selector		kiest welke individuen uit een populatie ouders worden	`rank`	de beste zoveel uit de populatie
variator	*mutator*	verandert een of meer elementen van een ouder	`network local`*	verplaatst sensoren, waarbij de kans op een doellocatie kleiner wordt naarmate hij verder van de bronlocatie afligt (uitgedrukt in netwerkmodelknopen)
	crossover	combineert twee ouders tot een kind	`n-point crossover`	knipt op n punten het gen van 2 ouders door en mengt de segmenten in de kinderen
replacer		bepaalt welke ouders en kinderen deel uitmaken van de volgende generatie	`generational + elitism`	vervanging van de ouderpopulatie door de kindpopulatie, waarbij een bepaalde fractie met de beste kandidaat-oplossingen onder de ouders behouden blijft (*elitism*) ten koste van de slechtste kandidaat-oplossingen onder de kinderen (mits deze slechter presteren dan de te behouden ouders)
terminator		bepaalt wanneer de evolutie moet worden afgebroken	`change`*	wanneer de relatieve verandering in prestatie van generatie op generatie gedurende een aantal generaties onder een voorgeschreven drempel blijft, wordt de evolutie stopgezet

* in dit project aan bestaande gebruikte functiebibliotheek toegevoegde functionaliteit

In dit rapport wordt daarom gepoogd om deze aanname te vermijden door te kiezen voor optimalisatiedoelstellingen die zo dicht mogelijk tegen het netwerk en de operatie daarvan aanliggen en die zo weinig mogelijk aannames bevatten. Tabel 3 geeft een overzicht van de gebruikte doelstellingen met hun belangrijkste kenmerken. Hiervan zijn de pakkans en detectietijd in de literatuur veelgebruikte doelstellingen en de overige in dit rapport ontwikkeld en uitgewerkt (maar het is niet uit te sluiten dat met name de voor de hand liggende eerste twee ook elders al zijn toegepast). Enkele van de opgenomen doelstellingen zijn niet afhankelijk van besmettingsscenario's, maar bevatten een beschrijving van de netwerkhydraulica over een periode van 24 uur. Deze lenen zich hiermee beter voor toepassing op discontinue metingen (monsternames) dan de overige optimalisatiedoelstellingen.

De laatste twee in de tabel opgenomen doelstellingen zijn niet geïmplementeerd in de in dit project ontwikkelde onderzoekssoftware. De optimalisatie op gemak van bronbepaling is op zichzelf al een zeer rekenintensieve activiteit gebleken. Daarom wordt het hier nog een niveau bovenuit stijgen op dit moment en met de hier gebruikte benadering niet als haalbaar beschouwd. In feite geldt iets vergelijkbaars voor de robuustheid m.b.t. veranderingen in de netwerkconfiguratie. Het aantal mogelijke veranderingen is bijzonder groot, en voor iedere verandering moet de complete set berekeningen die de basis vormt voor de volledige beschouwing van alle eerdere optimalisatiedoelstellingen worden uitgevoerd. Anders gezegd, de benodigde rekentijd is het aantal te beschouwen veranderingen in de netwerkconfiguratie maal de eerder geconstateerde voorbereidende rekentijd (enkele dagen voor het Proeftuinmodel). Op dit moment is deze doelstelling dus te complex voor volledige optimalisatie. De effecten van een verandering in de netwerkconfiguratie op het kunnen bepalen van een besmettingsbron zijn eerder geïllustreerd (Van Thienen en Vries, 2013). In

dat rapport is aangetoond dat een enkele foute afsluiterstand in een hydraulisch model voor een compleet verkeerde identificatie van het brongebied van een besmetting kan zorgen. Het ligt voor de hand dat optimalisatie van de sensorconfiguratie op het voorkomen van een dergelijke miswijzing waardevol is. Anderzijds zijn acties en protocollen om het optreden van foute afsluiterstanden in hydraulische modellen te voorkomen een alternatieve benadering om hetzelfde effect te verkrijgen, namelijk het kunnen vertrouwen op de uitkomsten van de berekeningen van het hydraulische model.

Tabel 3: Overzicht van toegepaste optimalisatiedoelstellingen met karakteristieken.

	doelstelling	*omschrijving*		*scenario's*	*geïmplementeerd*	*continu of discontinu*
1	Dekkingsgraad	Die set van sensoren die water ontvangt van een zo groot mogelijk deel van het netwerk	netwerklengte	nee	ja	c/d
			netwerkvolume		ja	c/d
			# aansluitingen		ja	c/d
			# consumenten		ja	c/d
1b	redundantie *n* in dekkingsgraad	Die set van sensoren die water ontvangt van een zo groot mogelijk deel van het netwerk, voor iedere locatie tegelijkertijd op minimaal *n* sensoren		nee	ja	c/d
2	Pakkans	Die set van sensoren waarvoor de kans op detectie van een willekeurige besmetting gemiddeld het grootst is		ja	ja	c
3	Detectietijd	Die set van sensoren waarvoor de tijd vanaf het begin van een willekeurige besmetting tot de detectie ervan het kleinst is		ja	ja	c
4	Gemak van bronbepaling	Die set van sensoren waarvoor de gemiddelde omvang van het gereconstrueerde brongebied zo gering mogelijk is.		ja	ja	c
5a	Robuustheid m.b.t. falen van sensoren: kunnen detecteren	Die set van sensoren waarvoor fout-negatieve signalen zo min mogelijk invloed hebben op het kunnen detecteren van anomalieën (pakkans)		ja	ja	c
5b	Robuustheid m.b.t. falen van sensoren: bronbepaling	Die set van sensoren waarvoor het fout-positieve en fout-negatieve signalen zo min mogelijk invloed hebben de bepaling van het potentiële brongebied van een besmetting (omvang en ligging brongebied)		ja	nee	c
6	Robuustheid m.b.t. veranderingen in de netwerkconfiguratie (afsluiters etc.)	Die set van sensoren waarvoor wijzigingen in de netwerkconfiguratie zo min mogelijk invloed hebben op het kunnen detecteren van anomalieën.		ja	nee	c

2.6 Uitgangspunten bij optimalisatiedoelstellingen

Voor de verschillende toegepaste optimalisatiedoelstellingen zijn diverse uitgangspunten van toepassing. Deze uitgangspunten worden hier beschreven. Specifieke gebruikte terminologie wordt toegelicht in Tabel 4.

Tabel 4: Definities en toelichtingen bij de uitgangspunten voor de verschillende optimalisatie-doelstellingen.

observatietijd/venster	tijdsperiode vanaf het begin van een besmetting tot een voorgeschreven aantal uren na het begin van de besmetting; in feite is dit de periode waarin detectie acceptabel is
kandidaat-sensornetwerk	deelverzameling van de verzameling potentiële sensorlocaties; hiervan wordt de prestatie bepaald in het optimalisatieproces
straftijd	in de optimalisatie veronderstelde tijdsduur tussen het begin van een besmetting en de eerste detectie ervan in het geval dat er geen detectie optreedt (buiten het blikveld van het sensornetwerk)

1. dekkings-graad
- Voor ieder kandidaat-sensornetwerk wordt bepaald uit welk deel van het netwerk gedurende (een deel van) het observatievenster water wordt bemeten.
- Voor dit 'geziene' gedeelte van het netwerk wordt de lengte, het volume, het aantal aansluitingen of het aantal consumenten bepaald.
- Deze parameter beschrijft de prestatie van het kandidaat-sensornetwerk, waarbij meer beter is.

1b. redundantie
- Bij het bovenstaande wordt vereist dat een bronlocatie door minimaal *n* sensoren tegelijk wordt bemeten (redundantie *n*).

2. pakkans
- Voor ieder kandidaat-sensornetwerk wordt voor iedere potentiële besmettingslocatie bepaald of een besmetting door minimaal 1 of *n* (redundantie) sensoren wordt gemeten binnen het observatievenster.
- De fractie van het aantal besmettingen dat worden wordt gedetecteerd, is de prestatie van het kandidaat-sensornetwerk (meer is beter).

3. detectietijd
- Voor ieder kandidaat-sensornetwerk wordt voor iedere potentiële besmettingslocatie en –tijdstip bepaald hoe lang het duurt voordat deze besmetting door een sensor wordt gedetecteerd (minder is beter).
- Sensoren die de besmetting niet detecteren binnen het voorgeschreven tijdsvenster vanaf het begin van de besmetting krijgen een straftijd toegewezen gelijk aan de maximale verblijftijd in het netwerk van het water op de betreffende besmettingslocatie en het betreffende tijdstip van besmetting.

4. gemak van bronbepaling
- Gedurende de observatietijd wordt voor iedere tijdstap het potentiële brongebied bepaald op basis van die sensoren die een positief signaal laten zien; wanneer geen enkele sensor een positief signaal laat zien, wordt verondersteld dat het hele netwerk het potentiële brongebied is.
- De minimale berekende omvang (in kilometers leidinglengte) van het potentiële brongebied van al deze tijdstappen voor een enkele besmettingslocatie wordt als bepaalde omvang voor deze locatie genomen.
- Het gemiddelde van de bepaalde omvangen voor alle besmettingslocaties wordt als prestatieparameter van het netwerk gebruikt.
- Bepaling van het brongebied vindt uitsluitend plaats op basis van positieve sensorbepalingen (zie bijlage II).

5a. robuustheid pakkans

- Er wordt alleen uitgegaan van fout-negatieve bepalingen, oftewel besmettingen die wel aanwezig zijn, maar in een dusdanig lage concentratie dat ze niet door de sensor worden opgepikt. De belangrijkste motivatie hiervoor is dat fout-negatieve bepalingen een risico vormen voor de klanten van drinkwaterbedrijven, terwijl fout-positieve bepalingen met name hinderlijk zijn voor het waterbedrijf (maar als het waterbedrijf op basis van een fout-positieve bepaling actie onderneemt eveneens voor de klant).
- De relatieve afname van de pakkans bij het falen van één of meerdere sensoren in het sensornetwerk wordt beschouwd. Deze wordt gedefinieerd ten opzichte van de pakkans in de onverstoorde situatie
- Hierbij wordt de relatieve afname gewogen naar de kans op optreden van de betreffende combinatie van juist-positieve, juist-negatieve en fout-negatieve bepalingen, zij bijlage I.
- Indien in de onverstoorde situatie geen detectie van een besmettingsscenario plaatsvindt, wordt dit scenario niet in het totaal meegenomen.

2.7 Besmettingsscenario's

Zoals aangegeven in Tabel 3 is het mogelijk of zelfs noodzakelijk voor verschillende optimalisatiedoelstellingen om besmettingsscenario's te definiëren. Bij de optimalisatie wordt uitgegaan van een verzameling van besmettingsscenario's die gezamenlijk representatief worden geacht voor de mogelijke bedreigingen van de waterkwaliteit die in het netwerk ingebracht kunnen worden. Deze verzameling is equivalent aan het zogenaamde *threat ensemble* zoals dit in TEVA-SPOT wordt toegepast.

In principe is het aantal combinaties van mogelijke locaties, tijdstippen en toedieningsduren oneindig. Door de ruimtelijke en tijdsdiscretisatie in het hydraulisch model waarop de optimalisatie wordt gebaseerd, ontstaat er een eindig, maar in praktische zin een meestal nog steeds ondoenlijk groot aantal mogelijkheden. Voor kleine modellen is de gemakkelijkste benadering de veronderstelling dat op ieder knooppunt van het netwerk met gelijke kans op een willekeurig tijdstip (iedere uitvoertijdstap van het hydraulische model) een besmetting op kan treden. Hierbij kunnen ook nog verschillende duren van de besmetting worden voorgeschreven, maar praktisch gezien is het aantal ook hier beperkt. De meest voor de hand liggende opties zijn een kortdurende (2 uur) en een langdurende (tot het eind van de simulatie) besmetting.

Voor grotere modellen ligt het voor de hand om het aantal potentiële besmettingslocaties terug te dringen door dicht bij elkaar gelegen locaties als equivalent te beschouwen. *Dicht bij elkaar gelegen* kan hier op verschillende manieren worden opgevat. Zo kan gekozen worden voor een goede geografische spreiding, maar een zinnigere strategie is goede spreiding over de lengte van het netwerk of de aansluitingen. Bijlage I beschrijft de strategieën die hiertoe in het onderzoek zijn toegepast.

Met enige kennis van de besmettingsmechanismen kan verdere verfijning worden aangebracht in de besmettingsscenario's. Zo kan een hogere dichtheid aan potentiële besmettingsbronnen worden voorgeschreven voor die delen van het netwerk die vaker storingen vertonen, aangezien het verhelpen van een storing een risico van inbreng van vreemd materiaal met zich meebrengt. Dit kan bijvoorbeeld op basis van USTORE-gegevens

voor cohorten. Deze benadering is niet gekozen in dit rapport, vanwege de benodigde tijdsinvestering voor het verkrijgen en inbrengen van leidingmateriaalgegevens in de modellen. Andere risicopunten zijn brandkranen en gebieden met geplande werkzaamheden. Voor deze gevallen kunnen eenvoudig extra potentiële besmettingsbronnen in het model worden gedefinieerd. Een voorbeeld hiervan wordt uitgewerkt in paragraaf 3.11.

2.8 Potentiële sensorlocaties

Voor de potentiële sensorlocaties geldt in principe hetzelfde als voor de potentiële besmettingslocaties: meer mogelijkheden betekent meer rekentijd. Ook hier kan deze worden beperkt door nabijgelegen locaties (in verschillende zin, zoals hierboven besproken) als equivalent te beschouwen. Een alternatieve benadering is om uit te gaan van praktische locaties waar alle voor de sensoren benodigde factoren aanwezig zijn (elektriciteit, onderdak, bereikbaarheid, etc.). In dit rapport wordt de eerste benadering gekozen; in praktijk-toepassingen zal de tweede benadering vaak het meest zinvol en het eenvoudigst zijn. Voor de optimalisatie komen zij echter op hetzelfde neer: slechts een deelverzameling van alle mogelijke locaties kan worden beschouwd als potentiële sensorlocatie. Let op dat in dit onderzoek in de meeste gevallen de verzameling potentiële bronlocaties en de potentiële sensorlocaties hetzelfde is.

2.9 Representatie van waterkwaliteitssensoren en –bepalingen

Het transport van stoffen wordt in de toegepaste benadering beschreven met behoud van massa, oftewel een op een bepaalde plaats toegediende concentratie van een bepaalde stof verandert in het netwerkmodel ten gevolge van verdunning. Voor de sensoren wordt echter een eenvoudige, binaire benadering gekozen: een stof is óf wel óf niet aanwezig. Hierbij kan eenvoudig een detectielimiet worden ingesteld. In de in dit rapport gepresenteerde berekeningen is deze op 0 gezet.

2.10 Aannames

De belangrijkste aannames in de beschreven benadering zijn de volgende:

1. De gebruikte modellen vormen een goede en representatieve weergave van het werkelijke distributienetwerk, zowel in fysieke zin alsook in de zin van het functioneren van het systeem.
2. De toegepaste besmettingsscenario's zijn representatief voor de bedreigingen voor de waterkwaliteit waaraan het netwerk kan worden blootgesteld.
3. Het genetisch algoritme leidt met de toegepaste instellingen tot een globaal optimum.

De eerste aanname wordt besproken in Hoofdstuk 5, de tweede wordt nader beschouwd in Hoofdstuk 4 en de laatste in Bijlage II. Wat dit punt betreft is er echter geen garantie. Maar ook als het globale optimum niet wordt gevonden, wordt er in ieder geval een lokaal optimum gevonden, en daarmee en configuratie die beter presteert dan de meeste alternatieven. Ook deze zal in de praktijk naar verwachting goed bruikbaar blijken.

2.11 Gebruikte software

Alle hydraulische berekeningen zijn uitgevoerd met EPANET 2.0 (Rossman, 2000). Het transport van verontreinigingen is uitgevoerd met EPANET-MSX (Shang et al., 2008). Om deze bestaande programma's heen is een analyse- en visualisatieschil ontwikkeld in Python (Python, 2014), genaamd *Contamination Source Toolkit*. De basis hiervan is beschreven in Van Thienen en Vries (2013). De ontwikkelde en gebruikte software wordt in meer detail beschreven in Bijlage III.

2.12 Ideeën voor verdere ontwikkeling

Bij de uitwerking van ideeën van dit project zijn diverse nieuwe ideeën ontstaan voor verdere verbetering van de methodiek en benadering. Deze zijn nog niet geïmplementeerd, maar aanbevolen wordt om deze aspecten in eventueel vervolgonderzoek wel mee te nemen. De belangrijkste worden hier kort beschreven.

Genetisch algoritme

De locaties voor cross-over in het oudergenoom worden in de huidige implementatie willekeurig gekozen. Een effectievere benadering lijkt om deze locaties te kiezen op basis van de netwerktopologie, bijvoorbeeld bij voorkeur locaties die nauw verbonden clusters van knopen met elkaar verbinden. Dit zal ook een slimme ordening van de locaties in het genoom vergen.

Potentiële besmettingslocaties

De selectie van de deelverzameling van knooppunten die als potentiële besmettingslocaties worden beschouwd kan gebeuren op basis van de daadwerkelijke faalkans van leidingen, welke uit USTORE-gegevens kan worden geschat. Overigens kunnen kritieke knopen nu reeds handmatig worden geselecteerd.

Criterium 4: Gemak van bronbepaling

Bij de optimalisatie op de minimale gemiddelde omvang van het gereconstrueerde potentiële brongebied, wordt in geval van het niet-detecteren van een besmetting het hele netwerk als potentieel brongebied beschouwd. Wanneer een non-detectie niet op een of andere manier in de berekening wordt meegenomen, leidt een dergelijke optimalisatie tot een dichte clustering van de sensoren, zodat een beperkt aantal besmettingen heel goed kan worden ingeperkt. De volgende ideeën drukken de gewenste doelstelling wellicht beter uit:

- Druk de gemiddelde omvang van het gereconstrueerde potentiële brongebied uit als fractie van de dekking van de betreffende sensorconfiguratie en minimaliseer deze fractie in de optimalisatie.
- Maak de ‘straflengte’ niet gelijk aan de totale netwerklengte maar aan het complement van de dekking van de betreffende sensorconfiguratie, oftewel de lengte van dat deel van het netwerk dat niet door de sensoren wordt gezien.

3 Resultaten

3.1 Inleiding

Na uitvoering van tests op een eenvoudig proefnetwerk (zie Bijlage II) is de ontwikkelde methodiek en software toegepast op een hydraulisch model van een bestaand distributienetwerk. Dit netwerk en deze toepassing worden beschreven in de volgende paragrafen. Hierbij is de doelstelling tweeledig. Enerzijds worden de verschillende optimalisatiedoelstellingen en resulterende sensorconfiguraties met elkaar vergeleken en wordt vastgesteld in welke mate deze strijdig zijn dan wel goed samen opgaan. Anderzijds wordt het presteren van een bestaand sensornetwerk in dit distributienet beschouwd aan de hand van de opgestelde criteria en wordt gekwantificeerd wat er van een dergelijk netwerk en grotere configuraties mag worden verwacht.

Er wordt steeds gesproken over criteria als pakkans en detectietijd, zoals besproken in het vorige hoofdstuk. Let op dat steeds de *gemiddelde* pakkans of de *gemiddelde* detectietijd (idem dito voor de andere doelstellingen) voor alle beschouwde besmettingsscenario's (locaties en tijdstippen) wordt bedoeld. Hierbij wordt de spreiding rondom het gemiddelde buiten beschouwing gelaten. Let op dat ook deze spreiding voor drinkwaterbedrijven een interessante parameter kan zijn. Aanbevolen wordt om dit aspect in eventueel vervolgonderzoek nader te onderzoeken.

3.2 Proefgebied en hydraulische modellen

Voor toepassing van de ontwikkelde methodiek is gekozen voor het deelbalansgebied Leeuwarden-Bergum van Vitens' *Proeftuin Noardburgum*, ook bekend staand als de *Vitens Innovation Playground*. Hiervan is een gedetailleerd hydraulisch model beschikbaar. Om de benodigde rekentijd voor de uitgevoerde optimalisaties te beperken, is uitgegaan van een, eerder gemaakte en toegepaste, verdunde versie van dit model (LW2700). Dit model en het verdunningsproces staan uitgebreid beschreven in Van Thienen en Vogelaar (2013). Figuur 4 geeft het verdunde netwerkmodel weer, Tabel 5 enkele karakteristieken van het uitgangsmodel en model LW2700. Het model kent 48663 aansluitingen, met gemiddeld 1,97 klant per aansluiting. Let op dat het reservoir De Greuns (in- en uitgaande pijl in Figuur 4) zich in dit verdunde model niet op dezelfde wijze gedraagt als in het oorspronkelijke model. Ten gevolge hiervan zijn optimale sensorlocaties in de omgeving van het reservoir mogelijk niet representatief voor de werkelijke situatie. Voor de algemene trends en relaties die in dit rapport worden onderzocht, heeft dit echter geen noemenswaardige consequenties.

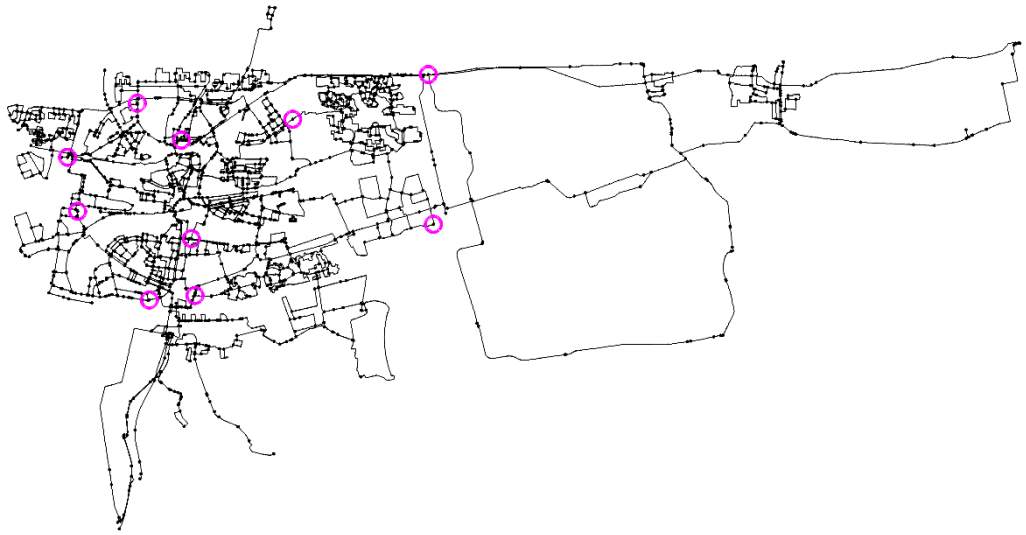
In het proefgebied zijn tien Eventlab-stations (zie De Graaf et al., 2012) opgesteld. Dit zijn generieke waterkwaliteitssensoren, die veranderingen in de brekingsindex van water met daarin opgeloste stoffen vaststelt. De locaties van deze stations zijn gekozen op basis van praktische overwegingen en geografische spreiding en zijn weergegeven in Figuur 5.



Figuur 4: Hydraulisch model LW2700 met in- en uitgaande waterstromen (rode pijlen).

Tabel 5: Vergelijking van het gebruikte hydraulische model LW2700 met het uitgangsmodel.

	uitgangsmodel	*LW2700*
aantal knopen	25091	2725
leidinginhoud (Ml)	24,93	23,01
leidinglengte (km)	612	306
aantal lussen	956	468

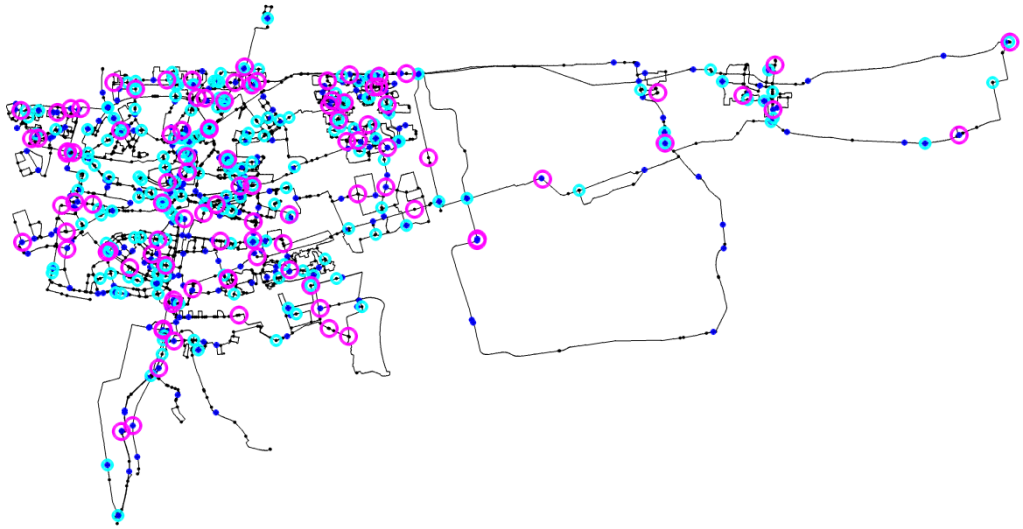


Figuur 5: Locaties van de 10 in het gebied geïnstalleerde Optiqua/Eventlab-stations, gekoppeld aan knopen in het verdunde netwerk.

3.3 Potentiële besmettings- en sensorlocaties

Om de benodigde rekentijd te beperken is gekozen om het aantal potentiële besmettings- en sensorlocaties te beperken, zie paragrafen 2.7 en 2.8. De gebruikte verzamelingen punten (100, 200 en 300 locaties) zijn weergegeven in Figuur 6. In de praktijk geldt een

vergelijkbare beperking, maar op andere gronden, bijvoorbeeld de eis van lokale toegankelijkheid of beschikbaarheid van elektriciteit. Ook zijn er enkele tests gedaan met alle knooppunten (zwarte bolletjes in Figuur 6) als potentiële locaties. Het effect van deze locatiekeuze wordt nader onderzocht in paragraaf 3.11. In de volgende paragrafen wordt in principe gewerkt met de basisselectie van 100 locaties, tenzij anders vermeld.



Figuur 6: Gebruikte verzamelingen uniform verdeelde potentiële besmettingslocaties, met n=100 (grote magenta cirkels), n=200 (kleine lichtblauwe cirkels) en n=300 (donkerblauwe stippen).

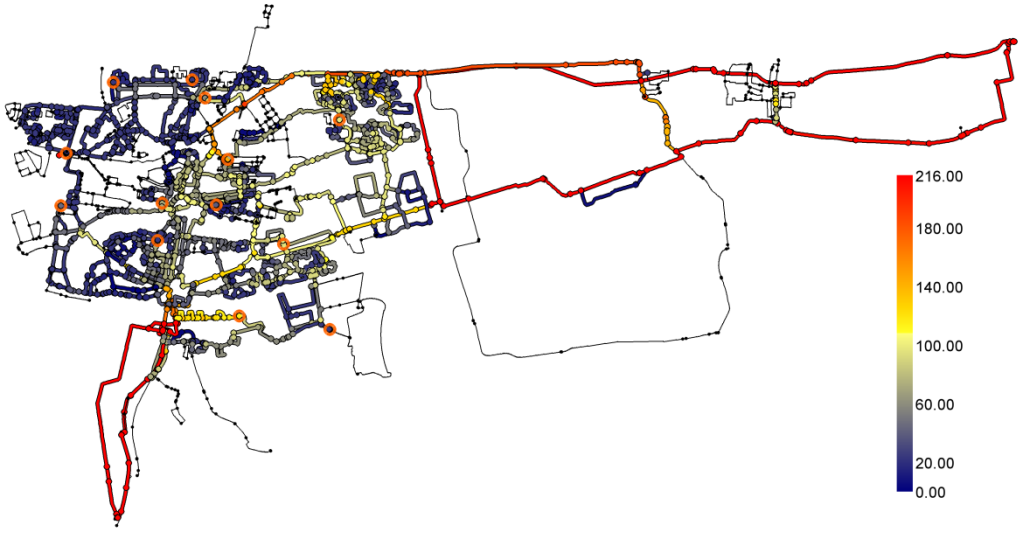
3.4 Criterium 1a: Dekkingsgraad

Voor de eerste doelstelling, maximale dekkingsgraad zonder redundantie, zijn optimale configuraties bepaald voor 10, 15 en 20 sensoren uit de potentiële verzameling van 100. De berekende prestaties voor deze optimale configuraties zijn weergegeven in Tabel 6, evenals die van de bestaande configuratie. Hierbij is steeds uitgegaan van maximale dekking in termen van *aantallen aansluitingen*. Let op dat aangezien in de modelberekeningen uitgegaan is van dezelfde verzameling potentiële locaties voor besmettingen en voor sensorplaatsing, er bij plaatsing van 100 sensoren in dit geval per definitie een dekkingsgraad van 1 is. De effecten van deze aanname worden nader beschouwd in paragraaf 3.11. Ter illustratie is de optimale configuratie voor 15 sensoren weergegeven in Figuur 7. Hierin zijn naast de locaties van de sensoren ook het geregistreerde/geziene aantal uurwaarden per netwerklocatie weergegeven. Dit houdt in dat per sensor ieder uur in een periode van 24 uur wordt bepaald of langs deze sensor water afkomstig van een willekeurige locatie in het netwerk stroomt. Per netwerklocatie wordt het aantal positieve uurwaarden voor alle sensoren gesommeerd weergegeven.

Twee zaken vallen op in Tabel 6: 1) de dekkingsgraad neemt voor dit netwerk maar langzaam toe met het toevoegen van meer sensoren, en 2) de bestaande configuratie presteert suboptimaal (60% van mogelijk prestatie voor dit aantal sensoren).

Tabel 6: Overzicht van berekende dekkingsgraad in termen van aantal aansluitingen voor optimale configuraties van 10, 15 en 20 sensoren (uit 100 potentiële locaties) voor redundantie n=1-3 (steeds geoptimaliseerd voor de betreffende redundantie) en de berekende dekkingsgraad van de bestaande configuratie van 10 sensoren.

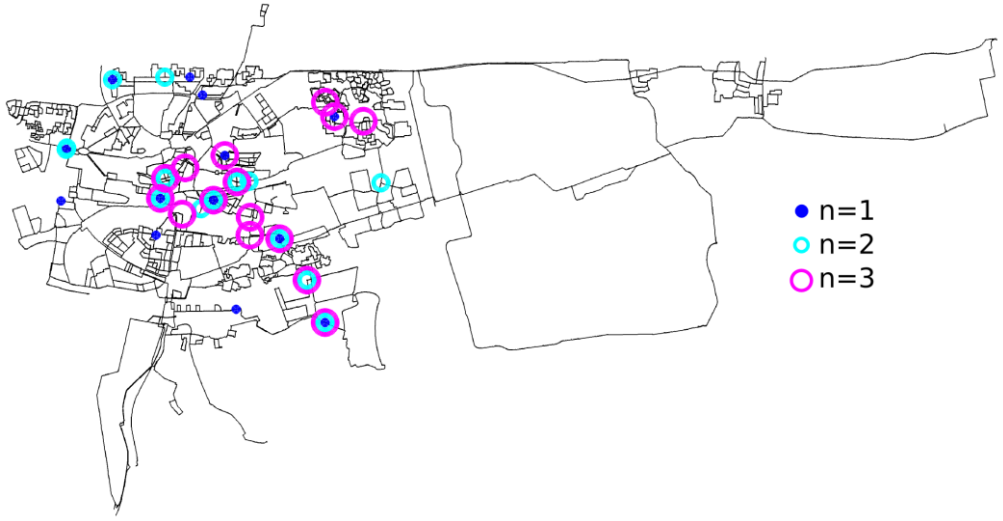
	10 sensoren	*10 sens. (bestaand)*	*15 sensoren*	*20 sensoren*
dekkingsgraad n=1	0,60	0,36	0,62	0,64
dekkingsgraad n=2	0,46	0,08	0,48	
dekkingsgraad n=3	0,29	0,02	0,31	



Figuur 7: Optimale locaties van 15 sensoren in de Proeftuin met als optimalisatiedoelstelling de dekkingsgraad (n=1) in termen van aantal aansluitingen. De kleurschaal geeft weer hoeveel uurmetingen er gesommeerd over alle sensoren water uit de verschillende delen van het netwerk zien.

3.5 Criterium 1b: Redundantie in de dekkingsgraad

Wanneer redundantie in de dekkingsgraad (nog steeds in de zin van aantallen aansluitingen) wordt vereist, heeft dit een duidelijke, en enigszins voor de hand liggende, invloed op de sensorconfiguratie: deze trekt samen, waardoor de vereiste redundantie wordt gerealiseerd, maar het totale dekkingsgebied kleiner wordt. Dit is duidelijk te zien in Figuur 8, waarin optimale configuraties voor 15 sensoren te zien zijn voor de situaties zonder (n=1) en met (n=2, n=3) vereiste redundantie. De bijbehorende dekkingsgraden zijn weergegeven in Tabel 6. Er is een snelle afname van de dekkingsgraad te zien voor toenemende vereiste redundantie. De in de Proeftuin bestaande configuratie verschaft nauwelijks redundantie.



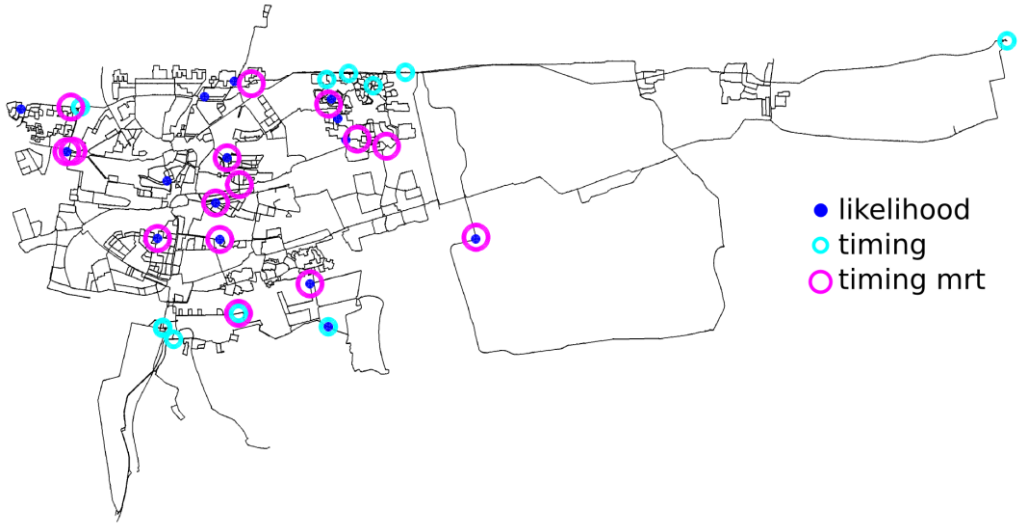
Figuur 8: Optimale sensorconfiguraties voor 15 sensoren met optimalisatie op maximale dekking in termen van aantallen aansluitingen zonder (n=1, blauwe stippen) en met redundantie (n=2, lichtblauwe cirkels, n=3, magenta cirkels).

3.6 Criterium 2: Pakkans

De pakkans is een in de literatuur veelgebruikt criterium (zie bijvoorbeeld Ostfeld et al., 2008) dat gemakkelijk te interpreteren is. Voor het testgebied is deze berekend voor optimale configuraties van respectievelijke 10, 15 en 20 sensoren, waarbij een duidelijke toename van de pakkans van 41% naar 55% te zien is van het eerste naar het laatste aantal, zie Tabel 7. Hier kan het toevoegen van extra sensoren dus duidelijk de moeite waard zijn. De pakkans van de bestaande configuratie is beduidend lager: ongeveer de helft van de maximaal haalbare waarde voor het aantal van 10 sensoren. De optimale configuratie voor 15 sensoren is ter illustratie weergegeven in Figuur 9.

Tabel 7: Pakkans (fractie gedetecteerde besmettingen ten opzichte van totaal aantal besmettingen) voor respectievelijk 10, 15 en 20 optimaal geplaatste sensoren en de bestaande configuratie van 10 sensoren.

	10 sensoren	*10 sens. (bestaand)*	*15 sensoren*	*20 sensoren*
pakkans (fractie aantal besmettingen)	0,41	0,22	0,50	0,55



Figuur 9: Optimale sensorconfiguraties voor 15 sensoren met optimalisatie op pakkans (likelihood, blauwe stippen), detectietijd met negeren van non-detecties (timing, lichtblauwe cirkels), en detectietijd met toepassing van de lokale maximale verblijftijd op de besmettingspunten voor non-detecties (timing mrt, grote magenta cirkels).

3.7 Criterium 3: Detectietijd

De detectietijd is eveneens een in de literatuur veelgebruikt criterium met een in principe heldere interpretatie. Echter, zoals in de vorige paragraaf is getoond, worden niet alle besmettingen gedetecteerd. Wanneer deze non-detecties worden genegeerd, wordt een optimale sensorconfiguratie gevonden waarin de sensoren zich dicht bij bronnen en transportleidingen bevinden, zie Figuur 9 voor het geval met 15 sensoren. Deze benadering is echter minder relevant voor de praktijk, omdat de niet-gedetecteerde besmettingen geen invloed hebben op het resultaat van de optimalisatie, maar wel op de klanten van het waterbedrijf. Wanneer de maximale verblijftijd op het punt en moment van besmetting als straftijd wordt toegekend aan niet-gedetecteerde besmettingen (zoals beschreven in paragraaf 2.6), worden alle besmettingen meegenomen in de beschouwing, maar dit ten koste van enige eenduidigheid in de betekenis van de gemiddelde detectietijd. Dit resulteert in een configuratie (voor 15 sensoren) die veel dichter bij de op pakkans geoptimaliseerde configuratie ligt (zie Figuur 9).

Tabel 8 toont de gemiddelde detectietijd voor hierop geoptimaliseerde configuraties van 10, 15 en 20 sensoren, respectievelijk, evenals die voor de bestaande configuratie van 10 sensoren. Opnieuw is de toegevoegde waarde van extra sensoren duidelijk: de gang van 10 naar 20 sensoren levert een reductie van de gemiddelde detectietijd van ongeveer 2½ uur of 24%. De bestaande configuratie heeft een beduidend langere gemiddelde detectietijd (3½ uur of 33%) dan een optimale configuratie met hetzelfde aantal sensoren.

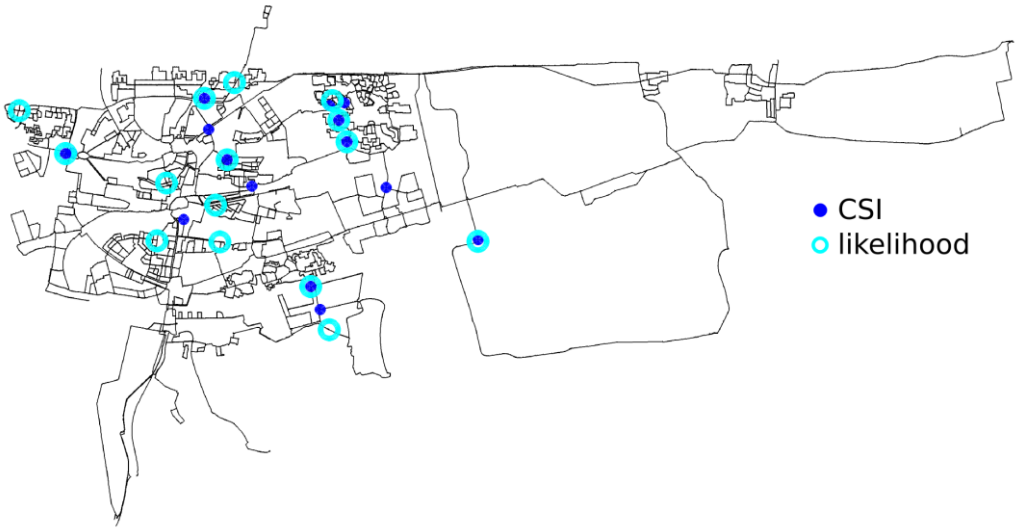
Tabel 8: Detectietijd (gemiddelde tijdsduur tussen het begin van een besmetting en de eerste detectie ervan) voor respectievelijk 10, 15 en 20 optimaal geplaatste sensoren en de bestaande configuratie van 10 sensoren.

	10 sensoren	*10 sens. (bestaand)*	*15 sensoren*	*20 sensoren*
detectietijd (uren)	10,43	13,89	9,13	7,96

3.8 Criterium 4: Gemak van bronbepaling

Optimalisatie van de configuratie van een populatie van 15 sensoren op optimaal gemak van bronbepaling resulteert in een configuratie die is weergegeven in Figuur 10. Ter vergelijking is ook de optimale configuratie m.b.t. pakkans weergegeven. Deze blijken kwalitatief redelijk vergelijkbaar. Ook hun prestaties vertonen overeenkomsten: de een presteert relatief goed op de andere doelstelling en andersom (dit wordt nader besproken in paragraaf 3.10).

Let op dat hier een afwijkende verzameling besmettingsscenario's is gebruikt: alle 100 potentiële besmettingslocaties (zie paragraaf 3.3) zijn beschouwd, maar er is slechts een besmettingstijdstip bekeken (middernacht) ten behoeve van het beperken van de benodigde rekentijd (zie paragraaf 2.7).



Figuur 10: Optimale sensorconfiguraties voor 15 sensoren met optimalisatie op gemak van bronbepaling (contamination source identifiability CSI, blauwe stippen) en pakkans (likelihood, lichtblauwe cirkels),

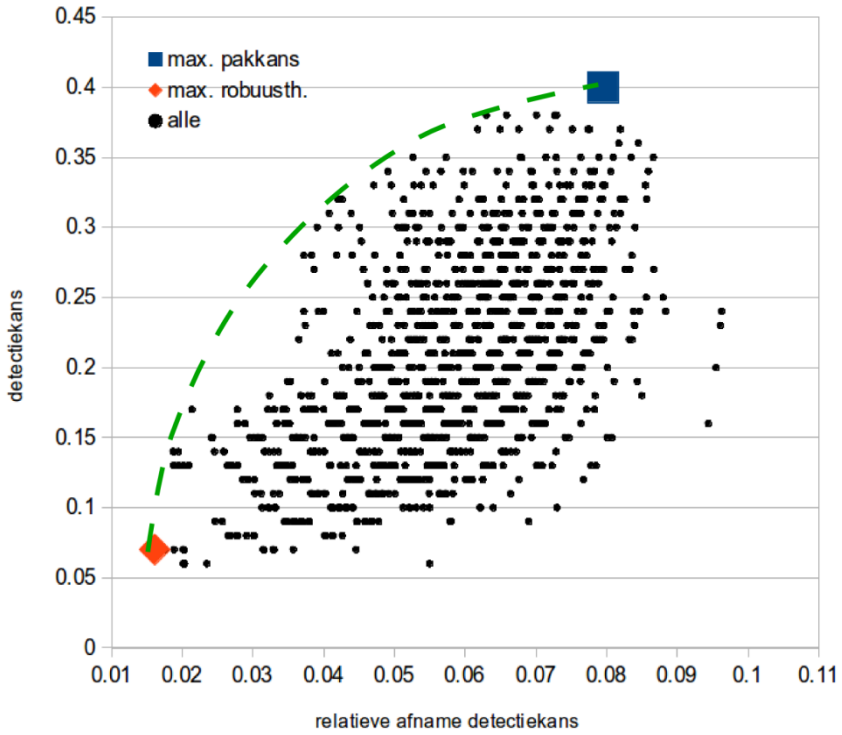
3.9 Criterium 5a: Robuustheid m.b.t. falen van sensoren: kunnen detecteren (pakkans)

Ook voor deze doelstelling is gerekend met een beperkte verzameling besmettingsscenario's (alle 100 locaties, slechts een tijdstip, zie vorige paragraaf) en is een optimalisatie uitgevoerd voor 10 sensoren. Hierbij is uitgegaan van een kans op fout-negatieve bepalingen van 10% per sensor. De prestaties van alle beschouwde configuraties op robuustheid (relatieve afname detectiekans) en pakkans zijn weergegeven in Figuur 11. Hierin zijn ook de configuratie (of een van de configuraties) met de beste robuustheid, oftewel de laagste relatieve afname van de detectiekans bij beschouwing van fout-negatieve bepalingen, en die met de hoogste detectiekans gemarkeerd. Let op dat dit optimalisatieproces niet op zoek is geweest naar de hoogste pakkans. Desalniettemin is de hier gevonden prestatie (0,40) vergelijkbaar met die van de in paragraaf 3.6 gevonden configuratie voor optimale pakkans.

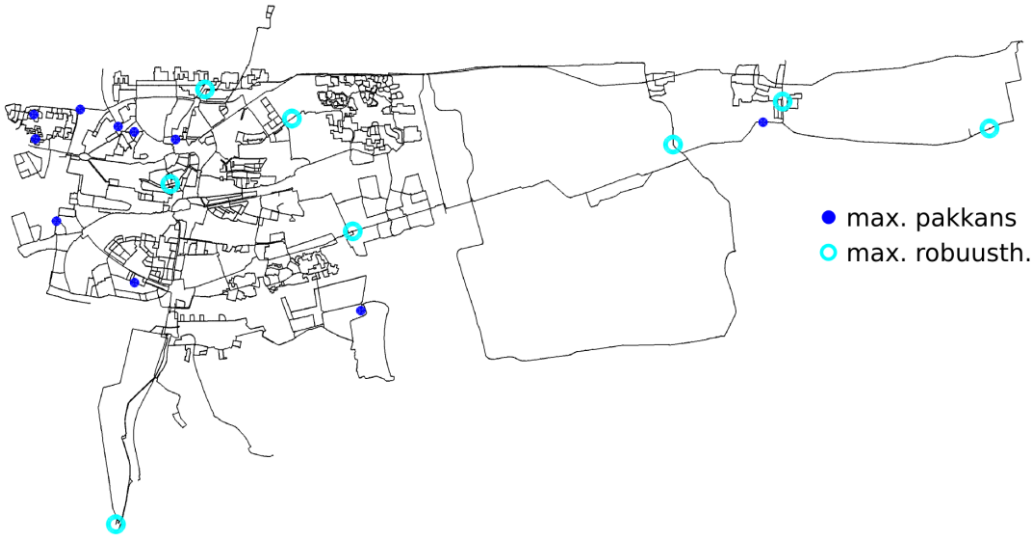
De bijbehorende sensorlocaties zijn weergegeven in Figuur 12. Let op dat de configuratie voor optimale robuustheid in twee gevallen twee sensoren op dezelfde locatie plaatst. Deze redundantie draagt vermoedelijk bij aan de robuustheid.

In de praktijk zal een goede balans tussen pakkans en robuustheid wenselijk zijn. Deze balans wordt gevonden op de linker bovengrens van de puntenwolk van Figuur 11, waar de

afweging tussen de twee criteria kan worden gemaakt. Let op dat de hier besproken optimalisatie geen formele multicriteriumoptimalisatie is. Wanneer deze zou worden uitgevoerd, zouden meer punten in dit grensgebied zijn geëvalueerd en de linker bovengrens van de puntenwolk duidelijker gedefinieerd zijn.



Figuur 11: Prestatie op robuustheid (relatieve afname detectiekans) en detectiekans voor alle door het optimalisatieproces beschouwde configuraties. De configuraties met de grootste robuustheid (laagste relatieve afname detectiekans) en de grootste detectiekans zijn gemarkeerd. Hiervoor zijn de sensorlocaties weergegeven in Figuur 12. De vermoedelijke ligging van het Paretofront is weergegeven met een groene onderbroken curve.



Figuur 12: Sensorconfiguraties voor maximale pakkans en maximale robuustheid zoals bepaald in de optimalisatie op robuustheid met betrekking tot het kunnen detecteren. De prestaties van de getoonde configuraties zijn weergegeven in Figuur 11.

3.10 Vergelijking van criteria

Van alle geoptimaliseerde configuraties voor de in paragrafen 3.3-3.8 besproken criteria voor 10 en voor 15 sensoren zijn de prestaties met betrekking tot al deze criteria berekend. Het doel hiervan is om te bekijken hoe een netwerk dat voor een doelstelling is geoptimaliseerd presteert op de andere doelstellingen. De resultaten van deze berekeningen zijn weergegeven in Tabel 9 voor 10 sensoren en in Tabel 10 voor 15 sensoren. Ook zijn de berekende prestaties van het bestaande Eventlab-netwerk (10 sensoren) en van een netwerktopologisch uniform verdeelde configuratie van 15 sensoren in de achtereenvolgende tabellen opgenomen.

Uit deze tabellen volgt een aantal observaties:

- In alle gevallen presteert een netwerk dat voor een bepaalde doelstelling is geoptimaliseerd beter dan alle andere beschouwde configuraties met hetzelfde aantal sensoren. Dit is natuurlijk de bedoeling, maar vormt een extra validatie van het optimalisatieproces en de gebruikte implementatie ervan.
- In alle beschouwde gevallen met 10 sensoren presteert de bestaande configuratie van Eventlab-stations slechter dan willekeurig welke geoptimaliseerde configuratie op alle doelstellingen.
- In alle beschouwde gevallen met 15 sensoren presteert de gevonden configuratie beter dan de uniforme configuratie. Hierop is één uitzondering: de op dekkingsgraad met redundantie n=3 geoptimaliseerde configuratie presteert slechter op dekking met redundantie n=1 dan de uniforme configuratie ([1] in Tabel 10). Aangezien de configuratie niet is geoptimaliseerd op n=3, is het wel denkbaar dat deze situatie zich voordoet.
- In alle beschouwde gevallen met 15 sensoren presteert de gevonden configuratie beter dan de corresponderende configuratie met 10 sensoren. Hierop zijn drie uitzonderingen, deze zijn gemarkeerd met [2] in Tabel 10. Ook dit is wel denkbaar, gezien in geen van de gevallen de optimalisatie heeft plaatsgevonden op het criterium waarvan de prestatie is bepaald.
- Pakkans en detectietijd lijken samen op te gaan, in de zin dat een netwerk dat is geoptimaliseerd op de ene doelstelling het ook relatief goed doet op de andere, en

andersom. In mindere mate geldt dit ook voor pakkans en gemak van bronbepaling. Een goede dekking (met name zonder redundantie) gaat juist slecht samen met een goede detectietijd.

Tabel 9: Overzicht van de prestaties van sensorconfiguraties van 10 sensoren die zijn geoptimaliseerd op pakkans (detection likelihood, DL), detectietijd (time to first detection, TFD), dekking en redundantie, en gemak van bronbepaling (contamination source identifiability, CSI). Ook zijn de berekende prestaties van de bestaande configuratie van Eventlab-stations in de tabel opgenomen.

		prestatie m.b.t. doelstelling, 10 sensoren					
	scenario	*pakkans*	*detectie-tijd (uren)*	*dekking en redundantie (fractie van alle aansluitingen)*			*CSI*
				n=1	*n=2*	*n=3*	
toegepaste sensor-configuratie	DL	0,41	11,31	0,52	0,39	0,19	263
	TFD	0,37	10,43	0,44	0,24	0,14	269
	CR1	0,33	12,67	0,60	0,30	0,13	284
	CR2	0,34	11,88	0,51	0,46	0,27	267
	CR3	0,30	11,54	0,37	0,32	0,29	268
	CSI	0,36	11,56	0,43	0,19	0,12	244
	bestaand	0,22	13,89	0,36	0,08	0,02	292

Tabel 10: Overzicht van de prestaties van sensorconfiguraties van 15 sensoren die zijn geoptimaliseerd op pakkans (detection likelihood, DL), detectietijd (time to first detection, TFD), dekking en redundantie, en gemak van bronbepaling (contamination source identifiability, CSI). Ook zijn de berekende prestaties van een niet geoptimaliseerde maar netwerktopologisch uniform verdeelde configuratie in de tabel opgenomen. [1]Geval waarin een geoptimaliseerde configuratie voor een andere doelstelling slechter presteert dan de uniforme configuratie; [2] gevallen waarin geoptimaliseerde configuraties met 15 sensoren slechter presteren op doelstellingen andere dan waarop is geoptimaliseerd dan de corresponderende configuratie met 10 sensoren.

		prestatie m.b.t. doelstelling, 15 sensoren					
	scenario	*pakkans*	*detectie-tijd (uren)*	*dekking en redundantie (fractie van alle aansluitingen)*			*CSI*
				n=1	*n=2*	*n=3*	
toegepaste sensor-configuratie	DL	0,50	9,22	0,54	0,34[2]	0,26	262
	TFD	0,45	9,13	0,52	0,40	0,22	272[2]
	CR1	0,39	11,80	0,62	0,33	0,21	272
	CR2	0,40	11,11	0,55	0,48	0,29	263
	CR3	0,38	10,28	0,41[1]	0,26[2]	0,31	268
	CSI	0,43	9,67	0,47	0,30	0,16	238
	uniform	0,26	12,48	0,44	0,25	0,11	285

Wanneer een waterbedrijf een configuratie wil bepalen die op meerdere doelstellingen zo goed mogelijk presteert, kan op eenvoudige wijze een gecombineerde prestatie-indicator worden opgesteld waarin de prestaties op individuele doelstellingen gewogen worden meegenomen. Om geen appels met peren te vergelijken, kan bijvoorbeeld de relatieve afwijking van een prestatie ten opzichte van de beste gevonden prestatie worden genomen in de gewenste richting. Een voorbeeld wordt gegeven in Tabel 11. Let op dat deze benadering eenvoudiger is dan een volledige multicriteriumoptimalisatie en waarschijnlijk niet tot de globale optimale oplossing leidt die deze multicriteriumoptimalisatie zou

opleveren. De configuratie is echter wel goed bruikbaar. Voor een volledige multicriterium-optimalisatie kan bijvoorbeeld gebruik worden gemaakt van het meer geavanceerde genetische algoritme NSGA-II (Deb et al., 2002).

Tabel 11: Voorbeeld van gewogen bepaling van beste configuratie voor meerdere doelstellingen. Individuele en gecombineerde optima zijn gemarkeerd met een grijze achtergrond.

	absoluut		absolute afwijking		relatieve afwijking		gewogen gemiddelde
	pak-kans	*detectie-tijd*	*pak-kans*	*detectie-tijd*	*pak-kans*	*detectie-tijd*	
weeg-factor					0,75	0,25	
DL	0,5	9,22	0	0,09	0,00	0,0099	0,0025
TFD	0,45	9,13	-0,05	0	0,10	0,00	0,075
CR1	0,39	11,8	-0,11	2,67	0,22	0,29	0,24
CR2	0,4	11,11	-0,1	1,98	0,20	0,22	0,20
CR3	0,38	10,28	-0,12	1,15	0,24	0,13	0,21
CSI	0,43	9,67	-0,07	0,54	0,14	0,059	0,12
uniform	0,26	12,48	-0,24	3,35	0,48	0,37	0,45

3.11 Effect van selectie potentiële locaties

Er zijn optimale configuraties bepaald voor verschillende combinaties van aantallen potentiële sensorlocaties en aantal potentiële bronlocaties om de effecten van deze verdunning te evalueren. Hiervoor is steeds het criterium dekkingsgraad met n=1 gebruikt, omdat deze doelstelling relatief snel is door te rekenen. De prestaties van deze optimale configuraties zijn weergegeven in Tabel 12, alle met betrekking tot de genoemde doelstelling.

Voor een aantal van 10 sensoren is er weinig verschil tussen varianten die zijn gebaseerd op 100, 200 of 300 potentiële locaties voor bron en sensor. Voor grotere aantallen sensoren zien we voor 300 locaties een trend die in het verlengde ligt van die voor 100 locaties. Een vergelijking met een berekening waarin alle netwerkknopen een potentiële bron- en sensorlocatie is laat veel grotere dekkingsgraden zien, met 7-15 procentpunt. Om dit verschil nader te duiden zijn er optimalisaties uitgevoerd met 100 potentiële bronlocaties en 2725 potentiële sensorlocaties en andersom. Het valt op dat met name de situatie met 100 potentiële sensorlocaties en 2725 bronlocaties slechter presteert; de situatie met 2725 bronlocaties en 100 potentiële sensorlocaties presteert vergelijkbaar met de “volledige” situatie (2725+2725 locaties). Hieruit kan worden geconcludeerd dat een relatief kleine sample uit alle knooppunten voldoende besmettingsscenario’s biedt om een representatief beeld te vormen, maar dat door verdunning van de potentiële sensorlocaties waardevolle locaties niet in de beschouwing worden meegenomen.

Ter verificatie is ook de dekkingsgraad van de bestaande configuratie van 10 Eventlab-stations met alle potentiële besmettingslocaties doorgerekend. Zoals op grond van de vorige alinea te verwachten valt, is de prestatie identiek aan de in paragraaf 3.4 besproken prestatie op basis van 100 potentiële besmettingslocaties.

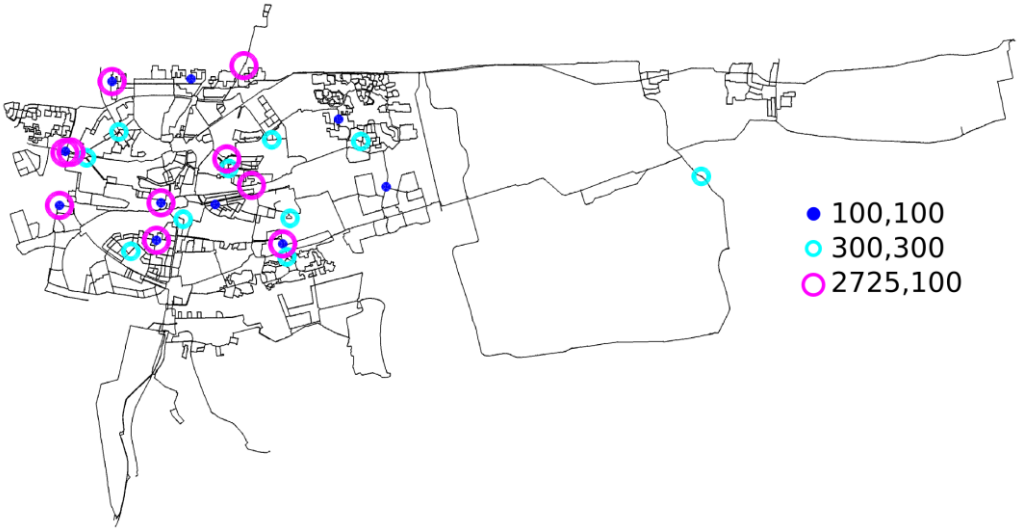
Figuur 13 toont drie optimale configuraties uit Tabel 12 (met * gemarkeerd) die (vrijwel) identiek presteren. Zoals te zien is, vertonen de eerste en de derde configuratie een redelijk mate van overlap, maar is de tweede volledig anders. Dit illustreert de reeds in eerder werk gedane constatering dat er veel verschillende configuraties kunnen bestaan die vergelijkbaar presteren. Naar verwachting zullen deze verschillende configuraties op andere doelstellingen echter niet noodzakelijkerwijs ook vergelijkbaar presteren.

De bevindingen van deze paragraaf hebben de volgende consequenties voor de interpretatie van de het voorgaande gevonden configuraties en berekende prestaties:

- Vergelijkingen moeten worden uitgevoerd tussen scenario's die gebaseerd zijn op dezelfde basisverzameling van potentiële sensorlocaties.
- Deze vergelijkingen zijn wel degelijk zinvol, omdat de uitgangssituatie steeds identiek is en de opschaling naar grotere aantallen potentiële sensorlocaties volgens voorspelbare trends lijkt te gaan.
- Extrapolatie van de geconstateerde trend leidt tot de conclusie dat ook het beschouwen van alle knooppunten van een verdund leidingnetmodel niet het volledige beeld schetst. Vermoedelijk is het wel afdoende om het onverdunde leidingnetmodel te beschouwen. Hierin vormen de knooppunten weliswaar eveneens slechts een deelverzameling van alle mogelijke sensorlocaties in het leidingnet, en de worden tussenliggende leidingen niet als potentiële locatie worden beschouwd, maar deze tussenliggende locaties zijn in feite hydraulisch gezien equivalent.
- Het beschouwen van een beperkte verzameling van potentiële sensorlocaties is in die zin wel een goede afspiegeling van de werkelijkheid, omdat daar ook de verzameling potentiële locaties beperkt is ten gevolge van praktische factoren, zoals de aanwezigheid van stroom.

*Tabel 12: Overzicht van berekende dekkingsgraad in termen van aantal aansluitingen voor optimale configuraties van 10 sensoren uit respectievelijk 100, 200, 300 en alle 2725 potentiële bron- en sensorlocaties met geselecteerde combinaties. De configuraties van de met een * gemarkeerde gevallen zijn weergegeven in Figuur 13.*

aantal bron-locaties	*aantal sensor-locaties*	aantal sensoren					
		10	15	20	30	50	100
100	100	0,60*	0,62	0,64			
200	200	0,63					
300	300	0,60*			0,68	0,73	0,77
2725	2725	0,67	0,71	0,79			0,87
100	2725	0,69	0,73	0,74	0,78		
2725	100	0,59*	0,63	0,63	0,64		
2725	bestaand	0,36					



Figuur 13: Optimale sensorlocaties met betrekking tot dekkingsgraad (n=1, aantallen aansluitingen) voor verschillende combinaties van aantal potentiële bronlocaties (eerste getal) en potentiële sensorlocaties (tweede getal). Deze gevallen zijn gemarkeerd in Tabel 11 met een asterisk.

4 Beschouwing en implementatie

4.1 Visie en doelstellingen

De resultaten die zijn gepresenteerd in het vorige hoofdstuk illustreren twee zaken heel duidelijk, namelijk dat 1) numerieke optimalisatie resulteert in een betere configuratie dan een evenredige verspreiding of plaatskeuze op basis van gevoel en ervaring, en 2) dat een sensornetwerk dat voor een bepaald doel wordt ingericht verre van optimaal kan functioneren wanneer het ook voor andere doelstelling wordt ingezet. Deze "tegenstrijdigheid van doelstellingen" is reeds eerder geconstateerd (Ostfeld et al., 2008). Maar ook voor de in dit rapport nieuw geïntroduceerde doelstellingen blijkt deze op te gaan; met name de tegenstelling tussen dekkingsgraad en detectietijd wordt opgemerkt. Het is daarom van groot belang dat een drinkwaterbedrijf goed nadenkt over de doelstelling van een eventueel te installeren sensornetwerk voor waterkwaliteit, voordat de configuratie hiervan wordt bedacht.

Een discussie en inventarisatie onder de leden van de projectgroep maakte duidelijk dat verschillende waterbedrijven wat dit betreft nog geen visie hebben ontwikkeld. Door de leden van de projectgroep wordt met name aan de volgende doelstellingen, gemiddeld genomen ongeveer in gelijke mate, een hogere prioriteit gegeven:

- dekkingsgraad;
- redundantie in dekkingsgraad;
- pakkans;
- detectietijd;
- representativiteit;
- gemak van bronbepaling.

4.2 Rekenwerk en mensenwerk

Rekentechnisch gezien leidt een optimalisatie in principe tot een optimale configuratie (met in paragraaf 2.10 besproken voorbehouden). Voorwaarde hierbij is uiteraard wel dat het hydraulisch model en de besmettingsscenario's waarop de optimalisatie is gebaseerd een goede weergave van de werkelijkheid vormen. Hier moeten de nodige kanttekeningen bij worden geplaatst, met handreikingen voor oplossingen hiervoor.

Ten eerste hebben de toegenomen mogelijkheden van hydraulische rekenpakketten enerzijds en de toegenomen beschikbaarheid van informatie over het netwerk geleid tot de zogenaamde *informatieparadox van het leidingnetmodelleren* (Mesman et al., 2014). Deze paradox kan als volgt worden beschreven: aangezien de netwerkanalist of –modelleur in de praktijk niet alles kan weten over de kwaliteit van informatie die wordt gebruikt in het opzetten van een hydraulisch model, hoe deze verwerkt wordt, welke marges de uitkomsten hebben en welke marges de uitkomsten mogen hebben, hebben de toegenomen mogelijkheden en beschikbare informatie eveneens geleid tot een toegenomen onzekerheid over de uitkomsten van de berekeningen. Dit is zeker geen diskwalificatie van hydraulische modellen, hun uitkomsten en hun gebruikers, maar wel een waarschuwing om het detailniveau van de berekeningen goed af te stemmen op het benodigde detailniveau van de uitkomsten van deze berekeningen.

Ten tweede geldt dat naast de kwaliteit van de gegevens waarop een netwerkmodel wordt gebaseerd, ook de actualiteit hiervan van groot belang is. De belangrijkste veranderingen in de configuratie van een distributienetwerk die op korte tijdschaal plaatsvinden zijn de manipulaties van afsluiters bij reparaties, spui-acties, etc. In de praktijk blijkt dat, ten gevolge van deze manipulaties en het niet altijd terugzetten in de oorspronkelijke stand, afsluiters met enige regelmatig niet staan zoals men denkt dat ze staan of zoals ze zouden moeten staan. Een enkele verkeerde afsluiterstand kan tot zeer afwijkende resultaten leiden bij bijvoorbeeld een backtraceberekening, waardoor voor een besmetting mogelijk het verkeerde potentiële brongebied wordt aangewezen (zie Van Thienen en Vries, 2013). Ook tijdelijke veranderingen in verbruikspatronen (weekeinden, vakanties, speciale evenementen zoals WK-finales, etc.) kunnen andere stromingspatronen in het netwerk tot gevolg hebben.

Voor besmettingsscenario's gaat men uit van de meest waarschijnlijk geachte manieren van besmetting van een netwerk. Per definitie wordt dus geen rekening gehouden met onverwachte scenario's, de zogenaamde zwarte zwanen met een kleine kans maar grote gevolgen (Taleb, 2007). Wel kan er rekening worden gehouden met de waarschijnlijkheid van optreden van een besmetting door bijvoorbeeld werkzaamheden, gietijzeren leidingen, of storingsfrequenties van het netwerk mee te wegen in de scenario's.

4.3 Verwachtingen met betrekking tot sensoring in distributienetwerken

Met de huidige beschikbare sensortechnieken en het bestaande wettelijke kader zijn de systemen die in drinkwaterdistributienetwerken (kunnen) worden toegepast realistisch gezien alle van de derde klasse (*detectie ter waarschuwing*), zoals gedefinieerd door Kroll en King (2010, zie ook paragraaf 1.2, criterium KK2). Zij kunnen een waarschuwing bieden, terwijl aanvullende analyses worden uitgevoerd, maar zijn op dit moment niet voldoende specifiek of nauwkeurig om in de tweede (*detectie ter bescherming*) of zelfs eerste (*detectie ter behandeling*) te vallen.

De in het vorige hoofdstuk opgesomde prestaties van optimale sensornetwerken m.b.t. verschillende doelstellingen illustreren hiernaast duidelijk dat het met een beperkt aantal sensoren onmogelijk is om een netwerk volledig te beschouwen en alle risico's af te dekken (dit nog afgezien van de feilbaarheid van systemen). Waterbedrijven doen er dan ook verstandig aan om met deze overweging rekening te houden in hun communicatie naar de buitenwereld.

4.4 Hoe verder te gaan met sensoring van waterkwaliteit in distributienetwerk

Met alle voorbehouden en kanttekeningen die hierboven zijn beschreven in gedachten, is er toch een duidelijke meerwaarde van enerzijds het plaatsen van waterkwaliteitssensoren in het distributienetwerk en anderzijds van de numerieke benadering van sensorplaatsing en optimalisatie hiervan te noemen.

Waterkwaliteitssensoren leveren hoe dan ook informatie op die *kan* leiden tot een snellere constatering van een besmetting en een meer doelgerichte omgang hiermee, zowel in de zin van het beperken van verdere verspreiding en het verwijderen als ook in de zin van de communicatie richting klanten. Het is echter belangrijk om de beperkingen van dergelijke systemen in gedachten te houden en de verwachtingen daarop aan te passen. Zolang sensoren en de installatie daarvan nog zodanig duur zijn dat het onmogelijk is om op iedere leiding een sensor te plaatsen, zal (bijna) volledige beveiliging van het netwerk ook onmogelijk blijven.

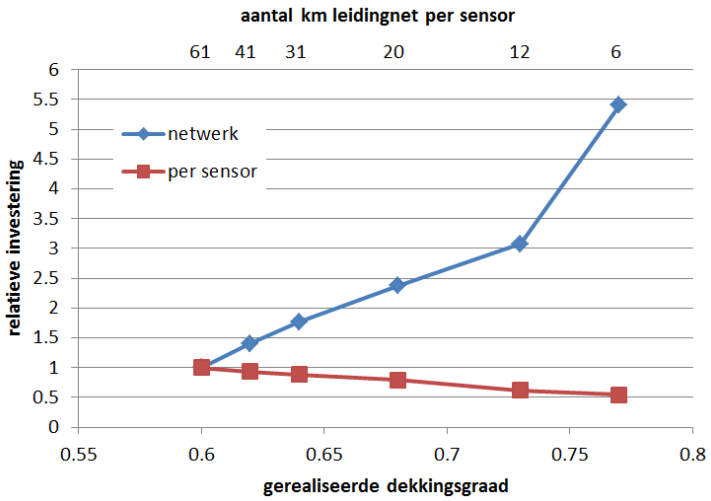
De gebruikte en ontwikkelde methodieken voor optimale locatiekeuze van waterkwaliteits-sensoren maken het mogelijk om verschillende doelstellingen tegen elkaar af te wegen en

tot een realistische verwachting van de opbrengsten van een dergelijk netwerk te komen. Dit draagt bij aan zowel de interne besluitvorming als de externe communicatie van beslissingen en implementaties op dit gebied. Voor “normale” risico’s (dus buiten de genoemde zwarte zwanen) kan het effect van sensoren worden gekwantificeerd, en kan ook de meerwaarde van het plaatsen van extra sensoren worden gekwantificeerd. Zo kan een afweging worden gemaakt tussen kosten en geboden mate van bescherming. Ook leveren de uitkomsten van deze berekeningen randvoorwaarden voor de responstijd die een bedrijf kan nastreven: bij een snelle detectie kan een snelle reactie veel blootstellingen van consumenten voorkomen, terwijl bij een langzame detectie of reactie van het waterbedrijf een respons veel minder effect zal hebben (zie ook Van Thienen en Vogelaar, 2013).

4.5 Naar een business case voor sensoring van waterkwaliteit

Op basis van de in dit rapport gepresenteerde prestatieberekeningen van een sensornetwerk als functie van het aantal optimaal geplaatste sensoren en de kosten van de aanleg en het onderhoud van het sensornetwerk kan een direct verband worden gelegd tussen de benodigde investering enerzijds en de opbrengsten in termen van de gekozen doelstelling van het netwerk anderzijds. Dit verband is de basis voor een business case voor een sensornetwerk, waarbij verder alleen nog de prestatie moet worden omgezet in een financiële opbrengst. Dit laatste punt is aan de waterbedrijven; voor het eerste wordt hier een voorbeeld uitgewerkt op basis van de aanschaf/installatie- en operationele kosten van Eventlab-stations van Optiqua.

Figuur 14 toont het berekende verband tussen de benodigde investering en de bereikte dekkingsgraad van het netwerk van het deelgebied Leeuwarden-Bergum van de Proeftuin Noordbergum. Hierbij zijn de operationele kosten meegenomen op basis van een gebruiksduur van 10 jaar. De totale kosten zijn weergegeven als veelvoud van de kosten voor een netwerk van 10 sensoren. De berekende curve toont in feite de competitie tussen enerzijds de afnemende kosten van de sensoren per stuk bij grotere aantallen (rode curve) en anderzijds het toenemende aantal extra sensoren dat nodig is om de dekkingsgraad verder te verhogen (zie Tabel 12). Dit resulteert in een knik in de berekende netwerkcurve bij een dekkingsgraad van 0,73. Links van deze knik is het relatief gemakkelijk en goedkoop om extra dekking te realiseren; grotere dekkingsgraden zijn moeilijker te realiseren (behoeven een groter aantal extra sensoren) en daarmee duurder.



Figuur 14: Benodigde relatieve investering (t.o.v. geval met 10 sensoren en een dekkingsgraad van 0,6) als functie van de gerealiseerde dekkingsgraad en de bijbehorende sensordichtheid.

De gemaakte berekening is slechts een voorbeeld. Voor andere doelstelling ziet zij er waarschijnlijk (enigszins) anders uit. Ook zal de vorm voor een van de Proeftuin afwijkende netwerktopologie anders zijn.

4.6 Aanbevelingen en randvoorwaarden

De toepassing van sensormetingen voor waterkwaliteitsbewaking steunt in belangrijke mate op hydraulische modellen, (idealiter) zowel in de ontwerpfase als in de interpretatie van hun puntmetingen in de context van een netwerk. Aanbevolen wordt om de kwaliteit van deze modellen zoveel mogelijk te waarborgen. Hiertoe kunnen de volgende stappen worden genomen:

- Opzet van een systeem voor registratie en regelmatige controle van afsluiterstanden in het leidingnet.
- Opzet of verbetering van een kwaliteitsborgingsmechanisme voor de gegevens-verzamelingen (zoals bijvoorbeeld in leidinginformatiesystemen) waarop hydraulische modellen worden gebaseerd.

Voorts wordt de waterbedrijven aanbevolen om, voor zover zij dit nog niet gedaan hebben, een visie te ontwikkelen met betrekking tot monitoring en bewaking van de waterkwaliteit in het distributienet met behulp van sensoren. Dit stelt hen beter in staat om nieuw op de markt komende waterkwaliteitssensoren voor hun eigen bedrijfsvoering op waarde te schatten en, indien gewenst, uiteindelijk te komen tot een doeltreffend sensornetwerk.

Ten slotte wordt aanbevolen om het ontwerp van eventueel te installeren sensornetwerk voor waterkwaliteitsbewaking te laten ondersteunen door numerieke evaluaties zoals in dit rapport beschreven. Hiermee kan de effectiviteit van het sensornetwerk significant worden verhoogd en beschikt het waterbedrijf bovendien over een middel om een business case op te stellen.

5 Conclusies en aanbevelingen

5.1 Conclusies

De in dit rapport uitgewerkte en geïmplementeerde methoden en hun toepassing op een leidingnetmodel van het deelbalansgebied Leeuwarden-Bergum in de Proeftuin Noordbergum hebben tot de volgende inzichten geleid:

- Numerieke optimalisatie resulteert duidelijk in een betere configuratie dan een evenredige verspreiding of plaatskeuze op basis van gevoel en ervaring (zie Tabel 13).
- Sensornetwerken die zijn geoptimaliseerd voor een bepaalde doelstelling presteren in de regel ook beter op andere doelstellingen dan configuraties met een evenredige verspreiding of plaatskeuze op basis van gevoel en ervaring (zie Tabel 13).
- Wanneer men uitgaat van een beperkte selectie potentiële sensorlocaties, worden optimale sensorconfiguraties gevonden die duidelijk minder goed presteren dan wanneer men alle netwerkknopen als potentiële sensorlocatie beschouwt. In de praktijk heeft men echter ook vanwege praktische randvoorwaarden te maken met een beperkte verzameling potentiële locaties.
- Ook met modelbeschouwingen die uitgaan van een beperkte verzameling potentiële sensorlocaties zijn trends duidelijk herkenbaar en vergelijkingen mogelijk en zinvol.

Tabel 13: Relatieve prestatie van geoptimaliseerde, huidig bestaande en uniforme sensornetwerken op verschillende doelstellingen in het deelgebied Leeuwarden-Bergum. De relatieve prestatie varieert van rood (de slechtste voor de betreffende doelstelling) tot groen (de beste voor de betreffende doelstelling). De beste en slechtste absolute waarde is steeds getalsmatig weergegeven. Steeds geeft een linkervakje de prestatie van de huidige configuratie van 10 Eventlab-stations in het deelgebied Leeuwarden-Bergum van de Proeftuin Noordbergum weer en een rechtervakje die van een uniform over het netwerk verspreide set van 15 sensoren. De afkortingen bij de optimale configuraties verwijzen naar de weergegeven doelstellingen onder prestaties. Pakkans en dekkingsgraad zijn weergegeven als fractie van het totaal, de detectietijd is in uren en bronbepaling geeft de gemiddelde omvang van het potentiële brongebied in kilometers weer.

(geoptimaliseer-de) configuratie	*prestaties met betrekking tot doelstellingen (10/15 sensoren)*											
	pakkans		*detectie-tijd*		*dekking en redundantie (fractie van alle aansluitingen)*						*bron-bepaling*	
					n=1		*n=2*		*n=3*			
pakkans	0,41	0,50										
detectietijd			10,4	9,13								
dekking n=1					0,60	0,62						
dekking n=2							0,46	0,48				
dekking n=3						0,41			0,29	0,31		
bronbepaling											244	238
huidig/uniform	0,22	0,26	13,9	12,5	0,36		0,08	0,25	0,02	0,11	292	285

prestatie	best										slechtst

- Een sensornetwerk dat voor een bepaald doel wordt ingericht kan verre van optimaal functioneren wanneer het ook voor andere doelstellingen wordt ingezet. Pakkans en detectietijd (en in mindere mate gemak van bronbepaling) zijn doelstellingen die, in ieder geval in het onderzochte netwerk, goed samen opgaan; een goede dekkingsgraad en een goede detectietijd lijken moeilijk samen te gaan (blijken tegenstrijdige doelen) voor een beperkt aantal sensoren (zie Tabel 13).
- Op basis van een eenvoudige weging kan worden bepaald welke sensorconfiguratie uit een reeks van voor verschillende doelstellingen uitgevoerde optimalisaties de beste balans voor de door een waterbedrijf belangrijk gevonden doelstellingen biedt.
- Met de huidige beschikbare sensortechnieken en het bestaande wettelijke kader zijn de systemen die in drinkwaterdistributienetwerken (kunnen) worden toegepast realistisch gezien alle van de derde klasse (*detectie ter waarschuwing*), zoals gedefinieerd door Kroll en King (2010). Zij kunnen een waarschuwing bieden, terwijl aanvullende analyses worden uitgevoerd, maar zijn op dit moment niet voldoende specifiek of nauwkeurig om in de tweede (*detectie ter bescherming*) of zelfs eerste (*detectie ter behandeling*) te vallen.

5.2 Aanbevelingen voor implementatie

Voor het succesvol opzetten, toepassen van en communiceren over een goed presterend sensornetwerk wordt aanbevolen de volgende overwegingen mee te nemen:

- Het is van belang dat een drinkwaterbedrijf goed nadenkt over de doelstelling van een eventueel te installeren sensornetwerk voor waterkwaliteit, voordat de configuratie hiervan wordt bedacht.
- Het is met een beperkt aantal sensoren onmogelijk om een netwerk volledig te beschouwen en alle risico's af te dekken (dit nog afgezien van de feilbaarheid van systemen).

Bij het opzetten van een netwerk kan een waterbedrijf goed gebruik maken van de hier besproken en ontwikkelde technieken, teneinde:

- het effect van sensoren te kwantificeren (ook de meerwaarde van het plaatsen van extra sensoren);
- verschillende doelstellingen tegen elkaar af te wegen en tot een realistische verwachting van de opbrengsten van een dergelijk netwerk te komen, ten behoeve van interne besluitvorming en externe communicatie ;
- een afweging te maken tussen kosten en geboden mate van bescherming;
- randvoorwaarden te leveren voor de responsstrategie en -tijd die een bedrijf kan nastreven.

Om de kwaliteit van hydraulische modellen, waarop zowel het ontwerp als de interpretatie van resultaten van een sensornetwerk kunnen worden gebaseerd, te waarborgen, wordt het volgende aanbevolen:

- Opzet van een systeem voor registratie en regelmatige controle van afsluiterstanden in het leidingnet.
- Opzet of verbetering van een kwaliteitsborgingsmechanisme voor de gegevens-verzamelingen (zoals bijvoorbeeld in leidinginformatiesystemen) waarop hydraulische modellen worden gebaseerd.

5.3 Aanbevelingen voor vervolgonderzoek en verdere ontwikkeling

- Voor een betere afweging van verschillende doelstellingen wordt aanbevolen om de besproken methoden in een multicriteriumoptimalisatie te beschouwen. Hiervoor bestaan reeds geschikte algoritmen, dus implementatie hiervan is relatief eenvoudig.
- Het belang van een goede samenhang tussen (het ontwerp van) sensornetwerken en responsstrategieën is reeds meerdere keren in het rapport opgemerkt. Verdere uitwerking van de behandelde of andere optimalisatiedoelstellingen met een verdergaande toespitsing op responsstrategie wordt aanbevolen.
- Voor toepassing van een sensornetwerk ten behoeve van monitoring van de bedrijfsvoering, bijvoorbeeld met behulp van *soft sensors*, wordt aanbevolen aanvullende strategieën voor optimale locatiekeuze op te stellen en te implementeren. Bij deze monitoring kan bijvoorbeeld worden gedacht aan het volgen van wettelijk voorgeschreven of voor de klant relevante waterkwaliteitsparameters in het distributienet, maar ook aan het bepalen van veranderingen van afsluiterstanden op basis van sensormetingen.
- De toepassing van netwerken van waterkwaliteitssensoren is veel gemakkelijker kostenefficiënt te maken, wanneer een sensornetwerk tegelijkertijd meerdere doelen kan dienen (Kroll en King, 2010). Een voorbeeld hiervan is de gecombineerde monitoring van wettelijk voorgeschreven of voor klanten relevante waterkwaliteitsparameters met bewaking op incidenten (besmettingen) in het distributienet. Aanbevolen wordt nader te onderzoeken hoe dergelijke combinaties kunnen worden ontworpen en hoe deze functioneren.
- Voor verbetering van het genetisch algoritme dat wordt gebruikt voor de optimalisatieberekeningen wordt aanbevolen een cross-overalgoritme te ontwerpen dat cross-overlocaties kiest op basis van de netwerktopologie.
- Voor de selectie van potentiële besmettingslocaties wordt aanbevolen de meerwaarde te onderzoeken van het gebruik van daadwerkelijke faalkansen van leidingen, welke uit USTORE-gegevens kunnen worden geschat.
- Aanbevolen wordt om een alternatieve benadering voor het meewegen van non-detecties bij optimalisatie van het gemak van bronbepaling te ontwikkelen waarbij het optimalisatieresultaat minder wordt gedomineerd door deze non-detecties.
- De huidige beschouwingen gaan altijd uit van een gemiddeld effect van een verzameling besmettingsscenario's. Aanbevolen wordt om in eventueel vervolgonderzoek ook de spreiding hiervan te beschouwen.

6 Verklarende woordenlijst

term	verklaring
besmettingsscenario	individuele combinatie van besmettingslocatie, tijdstip en duur; in de regel worden deze in verzamelingen beschouwd (bijvoorbeeld alle potentiële besmettingslocaties op alle hele uren van een dag voor een bepaalde duur)
dekkingsgraad	zie Tabel 3
detectietijd	zie Tabel 3
EDS	event detection systems; systemen voor continue monitoring die waarschuwen bij het constateren van afwijkingen
gemak van bronbepaling	zie Tabel 3
heuristiek	op ervaringen en probleemspecifieke inzichten gebaseerde oplossingstechniek
kandidaat-sensornetwerk	deelverzameling van de verzameling potentiële sensorlocaties; hiervan wordt de prestatie bepaald in het optimalisatieproces
metaheuristiek	op ervaringen en mogelijk probleemspecifieke maar generiek toepasbare inzichten gebaseerde oplossingstechniek zonder veronderstellingen over het op te lossen probleem
observatievenster	tijdsperiode vanaf het begin van een besmetting tot een voorgeschreven aantal uren na het begin van de besmetting; in feite is dit de periode waarin detectie acceptabel is (niet te lang duurt)
optimalisatie-algoritme	wiskundige methodiek om voor een op mathematisch-fysische wijze beschreven probleem een of meerdere optimale oplossingen te vinden
optimalisatiedoelstelling	de prestatieparameter die in de uitkomst van de optimalisatie zo klein of zo groot mogelijk moet zijn
pakkans	zie Tabel 3
Pareto-front	verzameling van deels optimale oplossingen van een optimalisatie met meerdere doelstellingen
redundantie n in dekkingsgraad	zie Tabel 3
robuustheid	mate waarin de prestatie ongevoelig is voor foute (in dit geval fout-negatieve) bepalingen
straftijd	in de optimalisatie veronderstelde tijdsduur tussen het begin van een besmetting en de eerste detectie ervan in het geval dat er geen detectie optreedt (buiten het blikveld van het sensornetwerk)
TEVA-SPOT	Threat Ensemble Vulnerability Assessment – Sensor Placement Optimization Tool (Berry et al., 2010); dit is software voor de bepaling van optimale sensorlocaties ten behoeve van bescherming van de bevolking tegen eventuele besmettingen

7 Literatuur

Berry, J., E. Boman, L. A. Riesen, W. E. Hart, C. A. Phillips, J.P. Watson, R. Murray (2010) User's Manual TEVA-SPOT Toolkit Version 2.4. Office of Research and Development, U.S. Environmental Protection Agency, EPA 600/R-08/041B.

Bush, C. A., en Uber, J. G. (1998). Sampling design methods for water distribution model calibration." Journal of Water Resources Planning and Management, 124(6), 334-344.

Deb, K.; Pratap, A.; Agarwal, S.; Meyarivan, T. (2002). "A fast and elitist multiobjective genetic algorithm: NSGA-II". IEEE Transactions on Evolutionary Computation 6 (2): 182. doi:10.1109/4235.996017

De Graaf B.R. Williamson F., Klein Koerkamp M., Verhoef J.W., Wuestman R., Bajema B., Trietsch E. and Van Delft W (2012) Implementation of an innovative sensor technology for effective online water quality monitoring in the distribution network. Water Practice & Technology, vol. 7, no. 4, (2012).

Jansen P, and Perez R (2011) Constrained Structural Design Optimization via a Parallel Augmented Lagrangian Particle Swarm Optimization Approach, International Journal of Computer and Structures, 89(13-14):1352–1366. http://dx.doi.org/10.1016/j.compstruc.2011.03.011

Kapelan, Z. S., Savic, D. A. en Walters, G. A. (2005). Optimal sampling design methodologies for water distribution model calibration." Journal of Hydraulic Engineering, 131(3), 190-200.

Kroll, D. en King, K. (2010) Methods for evaluating water distribution network early warning systems. Journal: American Water Works Association, 102(1), 1-11.

Mesman G., V. Sperber en P. van Thienen (2014) De informatieparadox van het leidingnetmodelleren. H2O-Online, 24 juni 2014.

Ostfeld, A. , James G. Uber, Elad Salomons, Jonathan W. Berry, William E. Hart, Cindy A. Phillips, Jean-Paul Watson, Gianluca Dorini, Philip Jonkergouw, Zoran Kapelan, Francesco di Pierro, Soon-Thiam Khu, Dragan Savic, Demetrios Eliades, Marios Polycarpou, Santosh R. Ghimire, Brian D. Barkdoll, Roberto Gueli, Jinhui J. Huang, Edward A. McBean, William James, Andreas Krause, Jure Leskovec, Shannon Isovitsch, Jianhua Xu, Carlos Guestrin, Jeanne VanBriesen, Mitchell Small, Paul Fischbeck, Ami Preis, Marco Propato, Olivier Piller, Gary B. Trachtman, Zheng Yi Wu en Tom Walski (2008) The Battle of the Water Sensor Networks (BWSN): A Design Challenge for Engineers and Algorithms, J. Water Resour.Plann. Manage., 134(6), 556-568.

Python (2014) http://www.python.org, laatst benaderd op 2 april 2014.

Rossman, L. A. (2000) EPANET 2 Users Manual. U.S. Environmental Protection Agency, report EPA/600/R-00/057.

Shang, F., Uber, J. G., Rossman, L. (2008) EPANET Multi Species Extension Software. U.S. Environmental Protection Agency ,Washington D.C., EPA/600/C-10/002.

Speight, V.L., Kalsbeek, W.D. en DiGiano, F.A. (2004) Randomized stratified sampling methodology for water quality in distribution systems. Journal of Water Resources Planning and Management, 130(4), 330-338.

Van Summeren, J. (2014) Numerieke validatie van backtracingberekeningen voor het bepalen van verontreinigingsbronnen in drinkwaterdistributiesystemen. BTO, in voorbereiding

Taleb, N.N. (2007) The Black Swan: The Impact of the Highly Improbable.

Van Thienen, P. en A. Vogelaar (2012) Optimalisatie en gevoeligheid van sensorplaatsing in het leidingnet. BTO 2012.017

Van Thienen, P. en D. Vries (2013) Backtracing van verontreinigingen in het distributienet - Methoden en bronbepaling met stochastische watervraag. BTO 2013.030

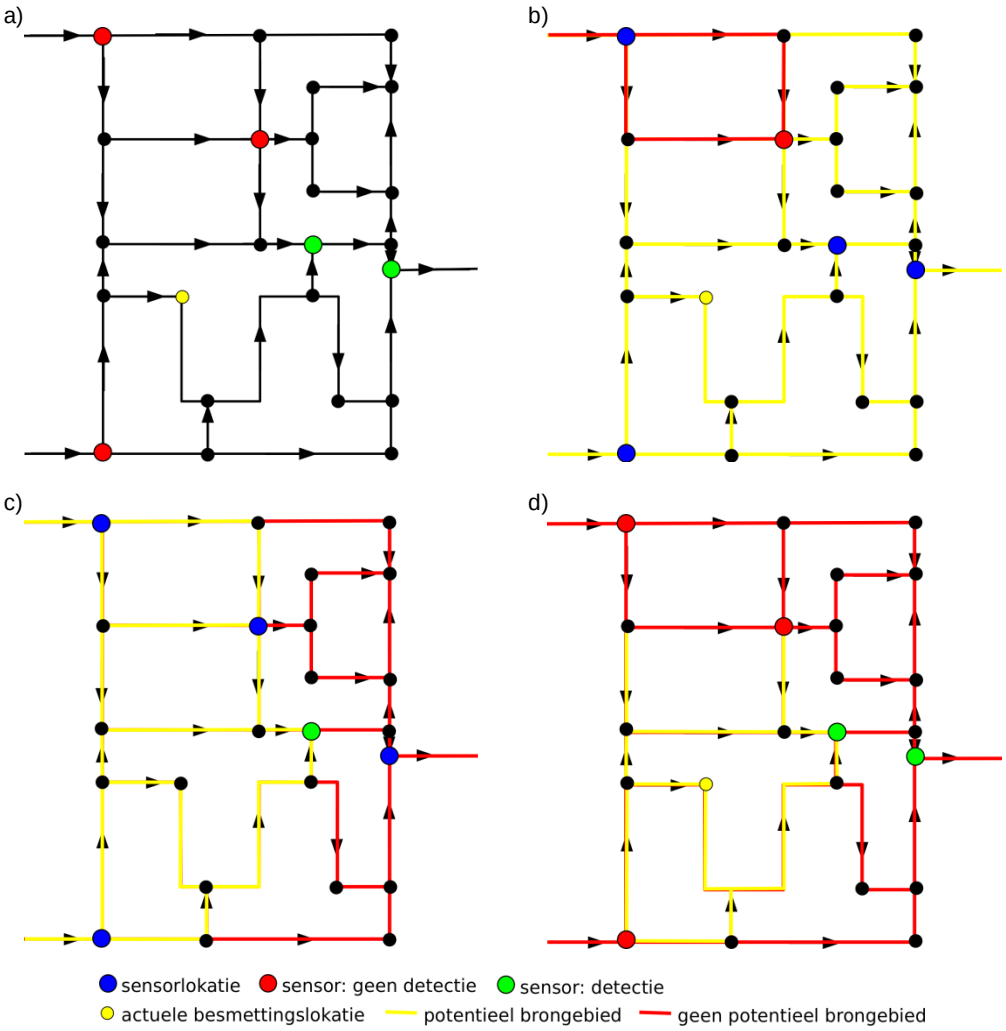
Van Thienen, P. en A. Vogelaar (2013) Optimale sensorplaatsing in deelgebied Leeuwarden-Bergum van de Proeftuin Noordbergum. BTO 2013.224(s).

Bijlage I: Algoritmen

Bepaling brongebied op basis van hydraulica en netwerkgeometrie en tijd

In beginsel wordt het potentiële brongebied *A* van een besmetting (Figuur 15a) eenvoudig bepaald door combinatie van de binaire herkomstkaarten, weergegeven als verzamelingen van knooppunten, voor de individuele sensoren die de besmetting oppikken S^+_i (zoals Figuur 15b) en de complementen (Booleaanse niet-operatie) van de binaire herkomstkaarten voor de individuele sensoren die de besmetting niet oppikken S^-_i (zoals Figuur 15c):

$$A = S^+_1 \cap S^+_2 \ldots \cap S^-_1 \cap S^-_2 \ldots$$



Figuur 15: Bepaling van potentieel brongebied voor een besmetting uitsluitend op basis van hydraulica en netwerk: a) besmettingslocatie (geel) en bijbehorende sensormetingen, b,c) potentiële herkomstkaarten voor individuele sensoren, d) gecombineerde potentiële herkomstkaart.

Het resulterende gecombineerde potentiële herkomstgebied *A* is weergegeven in Figuur 15d.

Met betrekking tot het gebruik van de negatieve bepalingen geldt een belangrijk voorbehoud: de aanname bij het beschreven gebruik is dat er voldoende tijd is verstreken sinds de aanvang van de (doorlopende) besmetting om de betreffende sensoren te bereiken. In specifieke situaties kan deze aanname gerechtvaardigd zijn. Voor algemene toepassing wordt uitgegaan van uitsluitend de positieve bepalingen (zo ook in dit rapport):

$$A = S^{+}_{1} \cap S^{+}_{2} \ldots$$

Let op dat de herkomstkaarten S^{+}_{i} zowel voor verschillende sensoren als voor gelijke en/of verschillende sensoren op verschillende tijdstippen gebruikt kunnen worden.

In het voorbeeld van Figuur 15 is gewerkt met een constant stromingsveld. In de praktijk varieert dit uiteraard door de dag heen. Wanneer deze variatie wordt verdisconteerd in de bepaling van de herkomstkaarten voor individuele sensoren, blijft de gepresenteerde benadering gelden. Beschouwing van stochastische variaties van dag tot dag vergt een probabilistische benadering met herhaalde simulaties met stochastisch gesimuleerde watervraag. Dit is uitsluitend van belang in de periferie van het tertiaire netwerk (zie Van Thienen en Vries, 2013).

Let op dat wanneer er geen sensor is die de besmetting oppikt, deze benadering met gebruikmaking van negatieve bepalingen zal resulteren in een potentieel brongebied dat de volledige periferie van het sensornetwerk omvat, oftewel alle knooppunten die hydraulisch gezien achter een sensor en niet voor een volgende sensor liggen.

Probabilistische bepaling van herkomstkaarten met faalkansen van sensoren

Voor iedere sensor kunnen, op basis van schattingen of meetgegevens, faalkansen worden gedefinieerd. Deze zijn in principe onafhankelijk voor fout-positieve en fout-negatieve bepalingen, omdat de mechanismen voor het optreden van fout-positieven respectievelijk fout-negatieven in de regel anders zullen zijn. De kans dat een bepaalde situatie zich voordoet (bijvoorbeeld sensor 1 functioneert en sensor 2 faalt), is in dit geval (aangezien het falen van de individuele sensoren onafhankelijk wordt beschouwd van het falen van de andere) gelijk aan het product van de kans dat de individuele componenten van die situatie zich voordoen (dus in dit voorbeeld [1 − faalkans] voor sensor 1 maal faalkans voor sensor2). Voor een geval met vier sensoren met elk verschillende faalkansen (volgens Tabel 14), kan de kans op het optreden van ieder van de mogelijke situaties worden uitgerekend, zoals weergegeven in Tabel 15.

Tabel 14: Kansen op het optreden van foute en correcte metingen voor individuele sensoren.

	observatie: positief		*observatie: negatief*	
	fout-positief	*juist-positief*	*fout-negatief*	*juist-negatief*
sensor 1	0.10	0.90	0.04	0.96
sensor 2	0.10	0.90	0.05	0.95
sensor 3	0.20	0.80	0.06	0.94
sensor 4	0.15	0.85	0.05	0.95

Tabel 15: Alle mogelijk optredende situaties en bijbehorende kansen op basis van de faalkansen van Tabel 14 voor een verzameling van 4 sensoren die respectievelijk 1,1,0,1 zouden moeten meten.

signaal	*falende sensoren*	*kans*
0,0,1,0	1,2,3,4	0.00009
0,0,1,1	1,2,3	0.00051
0,0,0,0	1,2,4	0.00141
0,1,1,0	1,3,4	0.00081
1,0,1,0	2,3,4	0.00081
0,0,0,1	1,2	0.00799
1,0,1,1	2,3	0.00459
1,1,1,0	3,4	0.00729
0,1,1,1	1,3	0.00459
0,1,0,0	1,4	0.01269
1,0,0,0	2,4	0.01269
0,1,0,1	1	0.07191
1,0,0,1	2	0.07191
1,1,1,1	3	0.04131
1,1,0,0	4	0.11421
1,1,0,1	geen	0.64719
	SOM	1.00000

Selectie van uniform verdeelde deelverzameling van knooppunten

Omdat het veel rekenkracht kost om alle mogelijke knooppunten als mogelijke besmettingsbron of sensorlocatie te beschouwen, en vele knooppunten door hun geografische of topologische nabijheid vrijwel equivalent zijn, is het zinvol om een representatieve deelverzameling van alle knooppunten te nemen. Het is echter wel wenselijk dat deze deelverzameling op een of andere wijze homogeen verdeeld is over het netwerk. Met behulp van het elders in dit rapport gebruikte genetische algoritme wordt een optimale spreiding verkregen op basis van de volgende criteria:

- buislengte;
- aantal aansluitingen;
- aantal consumenten;

Hiertoe worden eerst alle knooppunten van het netwerk geassocieerd met het dichtstbijzijnde knooppunt (in termen van aantal stappen door het netwerk) van de kandidaat-deelverzameling. Vervolgens wordt voor ieder van de zo ontstane clusters de prestatie (buislengte, aantal aansluitingen, etc.) berekend.

Bij de optimalisatie wordt de volgende L_2-norm geminimaliseerd:

$$L_2 = \sum (x_i - \mu)$$

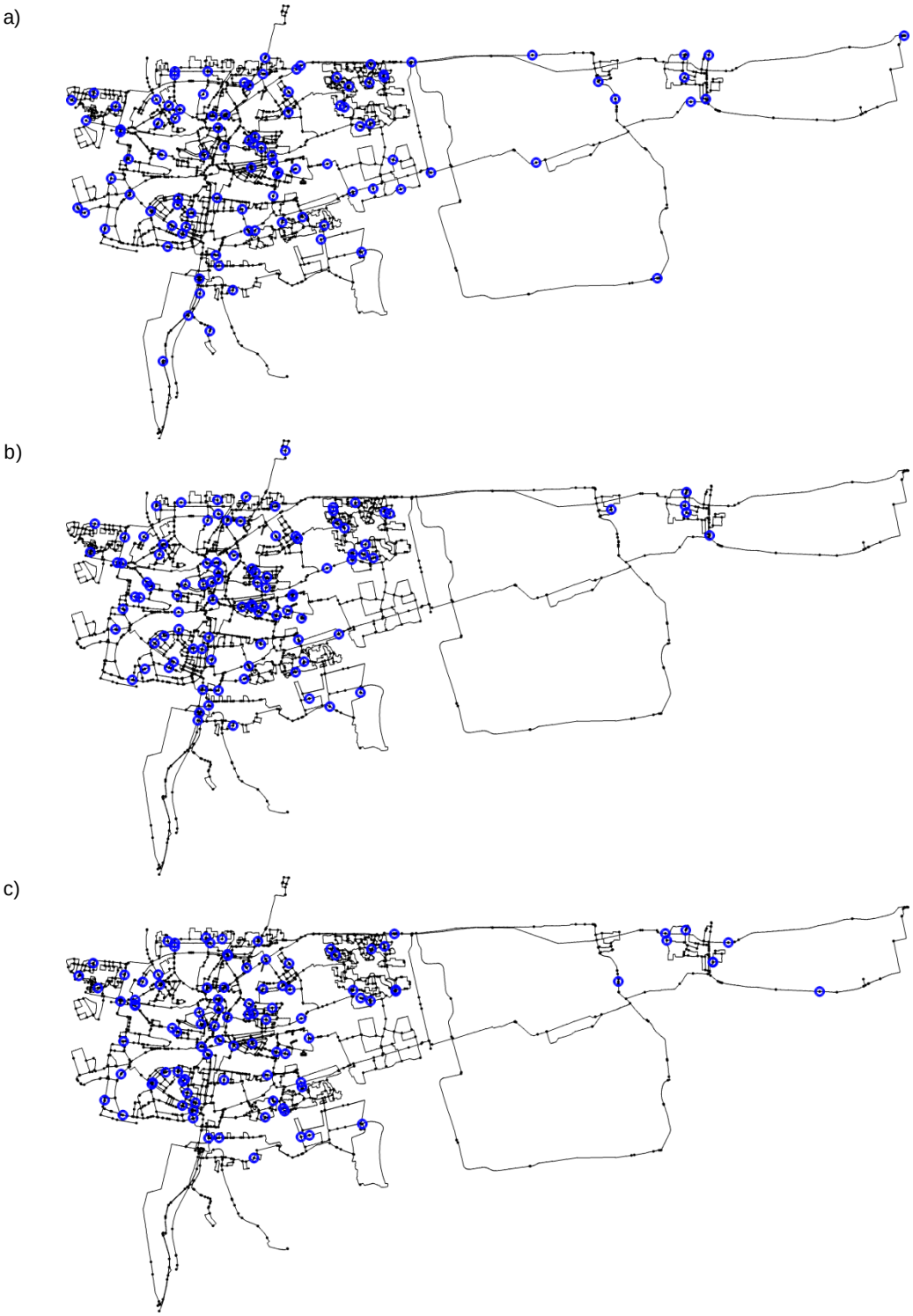
met x_i de waarden van de geselecteerde parameter en μ het gemiddelde van deze waarden. Het effect van deze minimalisatie is dat de gevonden oplossing een zo klein mogelijke spreiding in de gekozen parameter kent.

Voor elk van de drie criteria wordt een voorbeeld-deelverzameling van 100 knopen getoond in Figuur 16 (N.B. in hoofdstuk is een andere selectie van 100 knooppunten gebruikt, eveneens op basis van aantal aansluitingen). De deelverzamelingen voor aantallen aansluitingen en consumenten verschillen in de details. Deze verschillen zijn vermoedelijk in

eerste instantie toe te schrijven aan het genetisch algoritme – de verdelingen lijken min of meer equivalent. De deelverzameling geoptimaliseerd op buislengte toont een geografisch homogenere verdeling.

Twee punten dienen nog even genoemd te worden:

1. Ook hier is niet gegarandeerd dat het genetisch algoritme tot een globaal optimale oplossing komt. Er bestaan naar verwachting zeer veel min of meer equivalente configuraties.
2. De clusteringmethodiek die wordt gebruikt zorgt reeds voor een zekere vooringenomenheid in de selectie van individuele knooppunten voor de deelverzameling.



Figuur 16: Voorbeelden homogene spreiding van 100 knooppunten in een netwerk van circa 2700 knopen op basis van buislengte (a), aantal aansluitingen (b), en aantal consumenten (c).

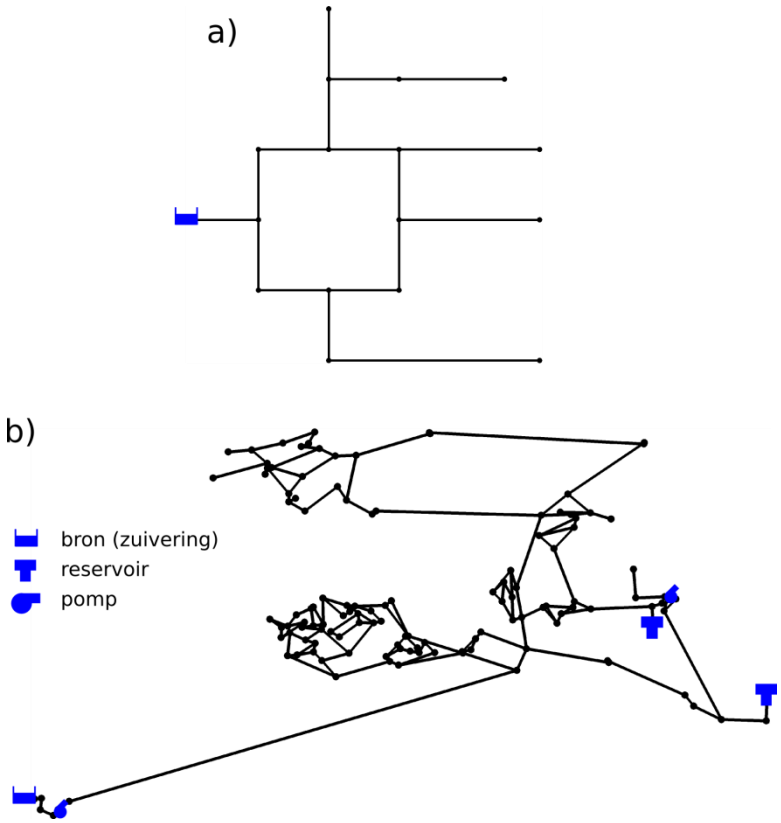
Bijlage II: Instellingen en evaluatie van optimalisatiedoelstellingen en -algoritmen

Benadering

Er zijn twee testnetwerken van verschillende grootte toegepast, waarin een optimale sensorplaatsing moet worden bepaald (zie Figuur 17). Het kleinste van de twee wordt gebruikt voor het testen en illustreren van de verschillende optimalisatiecriteria, gebruik makend van de brutekrachtbenadering om gegarandeerd optimale oplossingen te krijgen.

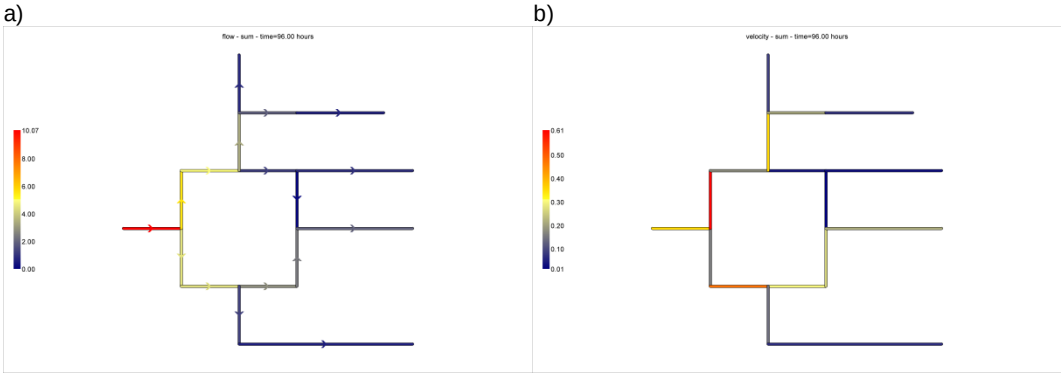
Het tweede netwerk wordt gebruikt om het functioneren en presteren van het genetisch algoritme te testen. Aangezien deze metaheuristieken bij een herhaling een ander resultaat zou kunnen opleveren, worden hiervoor de optimalisaties vijf keer uitgevoerd, en worden de gecombineerde resultaten van alle vijf weergegeven.

Testnetwerken



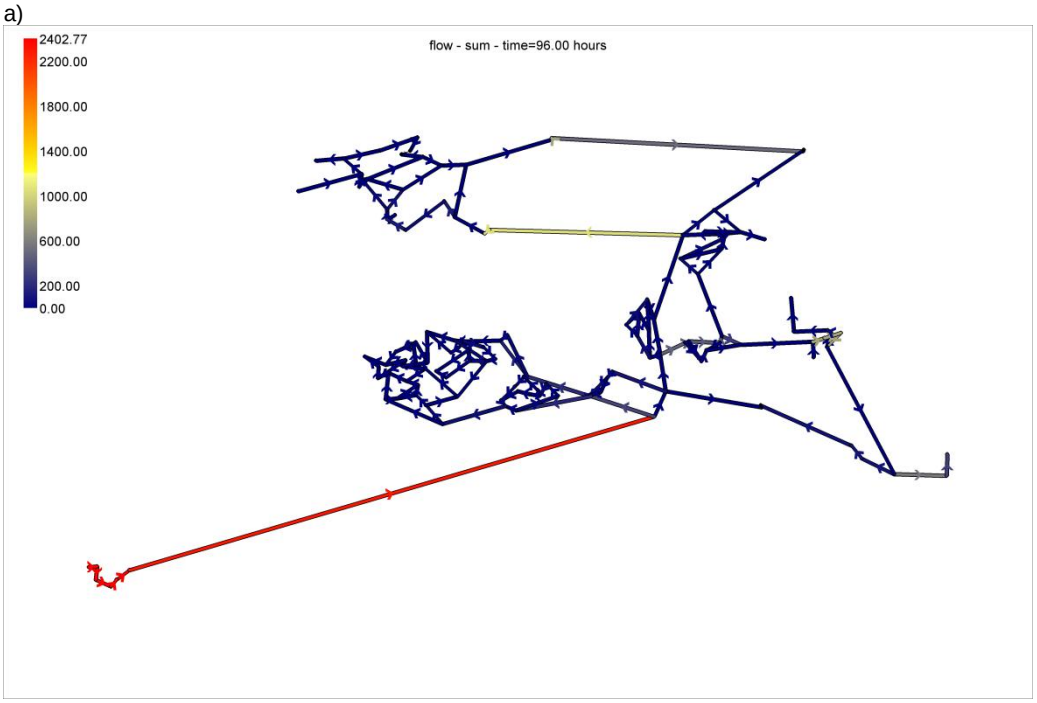
Figuur 17: Proefnetwerken die gebruikt zijn voor het testen en calibreren van het genetische algoritme. a) Eenvoudig netwerk met een bron en niet-variërende watervraag; b) complexer netwerk met variabele watervraag en sturingen (overgenomen van Ostfeld et al., 2008).

a) b)

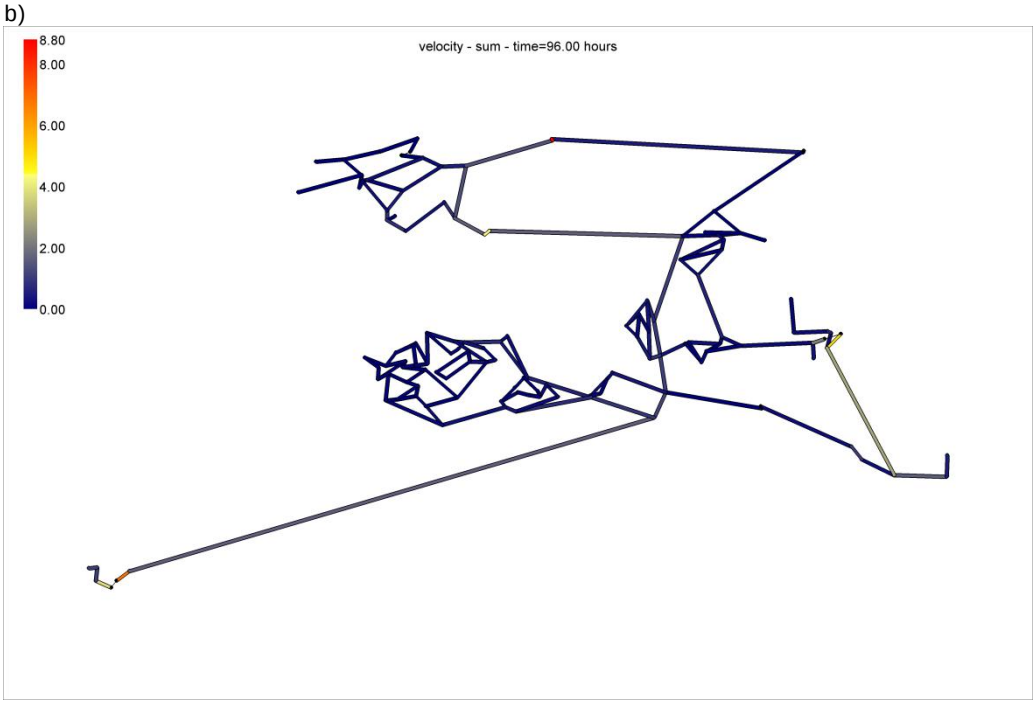


Figuur 18: Volumestromen (a) en stroomsnelheden (b) voor proefnetwerk 1. Gebruikte eenheden zijn m^3/uur resp. m/s.

a)



b)



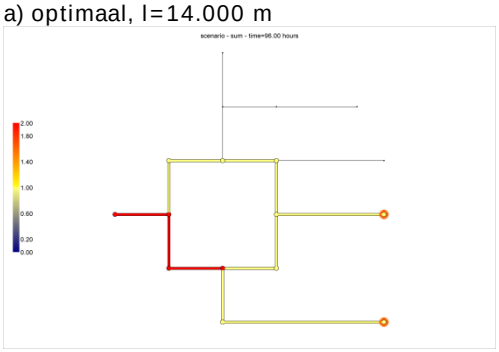
Figuur 19: Volumestromen (a) en stroomsnelheden (b) voor proefnetwerk 2 op tijdstip t=96 uur. Gebruikte eenheden zijn m^3/uur resp. m/s.

Tests van optimalisatiedoelstellingen

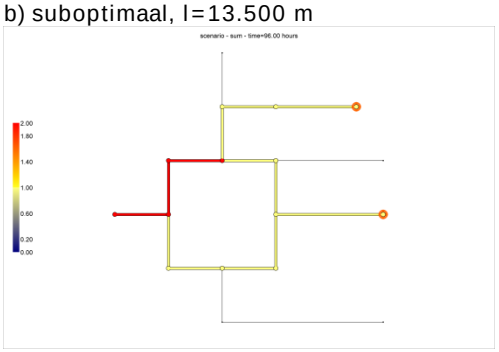
Criterium 1a: Dekkingsgraad

De dekkingsgraad wordt hier gedefinieerd als de buislengte (of het buisvolume, of de aan deze buizen gekoppelde aansluitingen of consumenten) waaruit water door een sensor wordt gezien, hetzij voor een bepaald tijdstip, hetzij gemiddeld voor een periode van 24 uur. Voor het niet in de tijd variërende stromingsveld van proefnetwerk 1a is dit simpelweg de leidinglengte (of buisvolume) vanaf de bron tot aan de sensor. Voor meerdere sensoren wordt het totaal zo groot mogelijk door de overlap te minimaliseren. Figuur 20 geeft een overzicht van de optimale en enkele suboptimale configuraties (geziene delen van het netwerk in geel en rood).

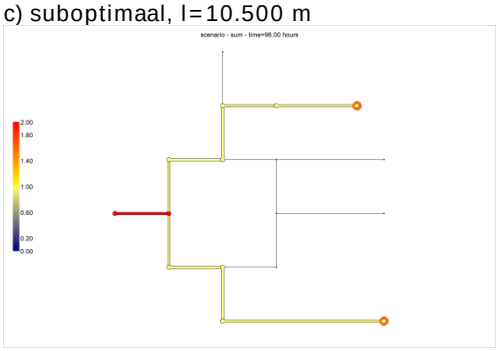
a) optimaal, l=14.000 m



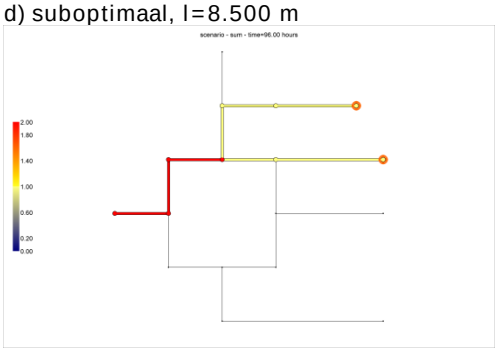
b) suboptimaal, l=13.500 m



c) suboptimaal, l=10.500 m



d) suboptimaal, l=8.500 m



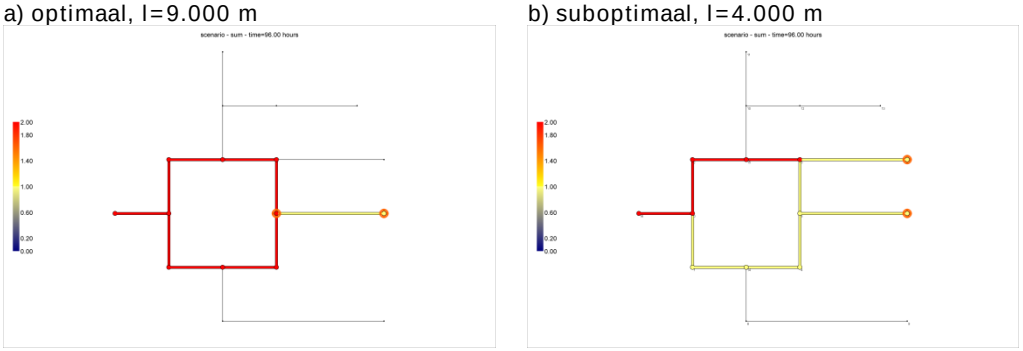
Figuur 20: Optimale en suboptimale sensorconfiguraties (oranje cirkels) voor maximale dekking van de sensoren zonder redundantie, met lengtes van het geobserveerde deel van het leidingnet, in geel en rood weergegeven

Criterium 1b: Redundantie in de dekkingsgraad

Vanuit het oogpunt van mogelijk falen van sensoren kan enige redundantie in het sensornetwerk wenselijk zijn. Ze kan bijvoorbeeld worden geoptimaliseerd op de dekkingsgraad met een redundantie van 2, hetgeen wil zeggen dat die configuratie wordt gezocht die een zo lang mogelijk deel van het netwerk (of zo groot mogelijk buisvolume) oplevert dat door twee sensoren wordt gezien. Ook hierbij geldt dat de beschouwing voor een bepaald tijdstip of gemiddeld voor een periode van 24 uur kan worden uitgevoerd. Figuur 21 geeft opnieuw een overzicht van de optimale en enkele suboptimale configuraties (redundant geziene delen van het netwerk in rood).

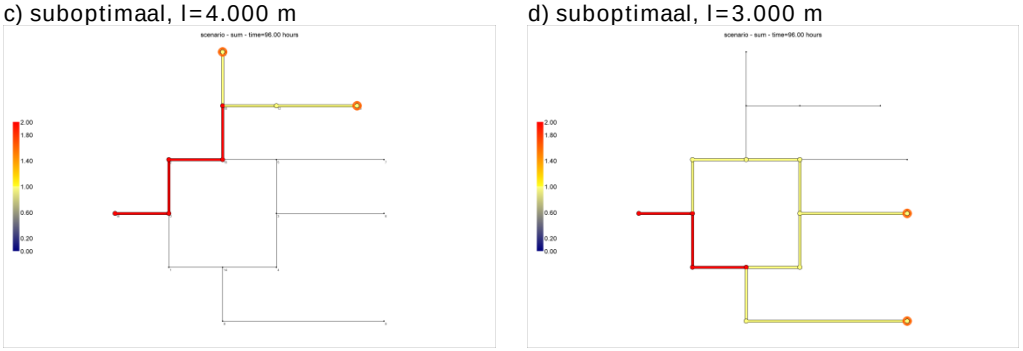
a) optimaal, l=9.000 m

b) suboptimaal, l=4.000 m



c) suboptimaal, l=4.000 m

d) suboptimaal, l=3.000 m



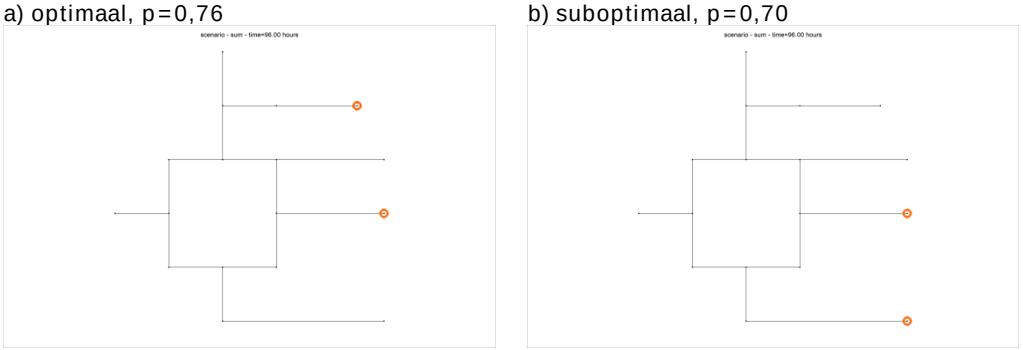
Figuur 21: Optimale en suboptimale sensorconfiguraties (oranje cirkels) voor redundantie n=2, met lengtes van het redundant geobserveerde deel van het leidingnet, in rood weergegeven.

Criterium 2: Pakkans

Voor het bepalen van de pakkans wordt een besmettingsscenario opgesteld. In dit geval veronderstellen we dat een verontreiniging tussen 24 en 28 uur na start van de simulatie op een willekeurig punt in het netwerk wordt ingebracht. Geëist wordt dat de verontreiniging binnen 24 uur na het begin ervan wordt opgemerkt door een sensor. De fractie van het aantal verontreinigingslocaties waarvoor de verontreiniging wordt opgepikt, wordt in de optimalisatie gemaximaliseerd. Omdat alle knooppunten een potentiële besmettingslocatie zijn, wordt in zekere zin geoptimaliseerd op het aantal knooppunten tussen de bron en de sensorlocatie. Figuur 22 geeft een overzicht van de optimale en enkele suboptimale configuraties met bijbehorende pakkans. De pakkans is in dit geïdealiseerde voorbeeld gelijk aan de het aantal knopen dat vóór minimaal één van de sensoren ligt gedeeld door het totaal aantal knopen.

a) optimaal, p=0,76

b) suboptimaal, p=0,70



c) suboptimaal, p=0,60

d) suboptimaal, p=0,50

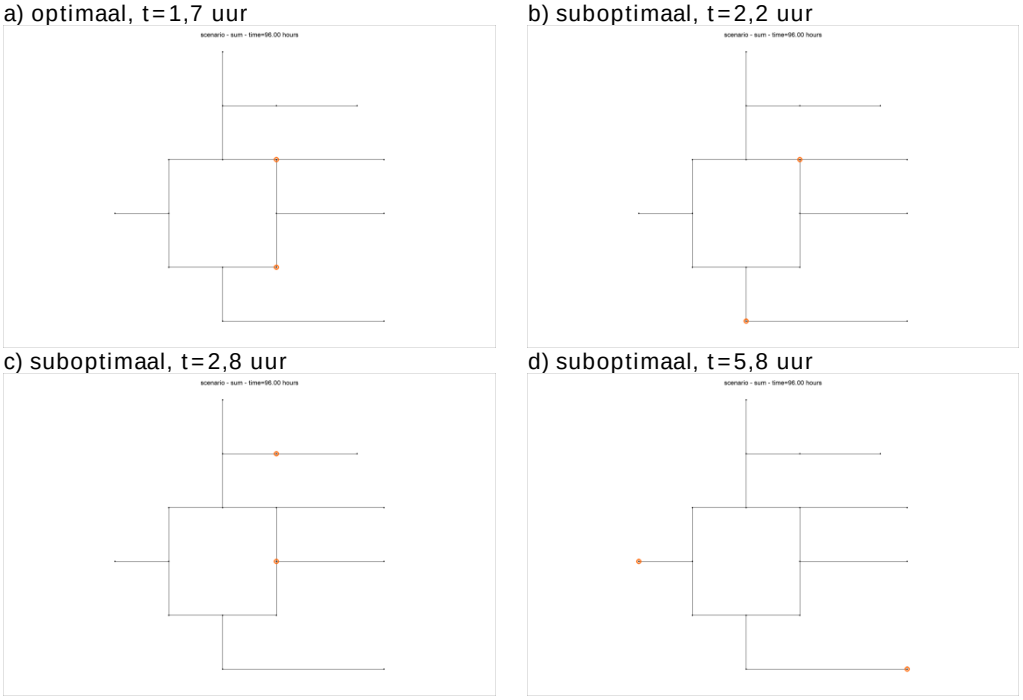
Figuur 22: Optimale en suboptimale sensorconfiguraties (oranje cirkels) voor de gemiddelde pakkans.

Criterium 3: Detectietijd

De detectietijd definiëren we hier als de tijdsduur tussen de aanvang van een verontreiniging en de eerste detectie ervan door een sensor. Mogelijk worden echter niet alle verontreinigingen door een sensor opgepikt (binnen de maximale tijdsduur of überhaupt). Aan deze gevallen wordt een zogenaamde straftijd toegekend. Wanneer optimalisatie van het sensornetwerk een zo klein mogelijke gemiddelde detectietijd nastreeft, blijkt de straftijd van grote invloed op de optimale sensorlocaties. Daarom wordt deze op een betekenisvolle netwerkparameter gebaseerd, namelijk de maximale verblijftijd voor het knooppunt en tijdstip van de besmetting uit het betreffende scenario. Hiermee worden punten aan het begin en aan het eind van het netwerk minder aantrekkelijk als sensorlocatie, de eerstgenoemde vanwege de lange resterende maximale verblijftijd en de laatstgenoemde vanwege de lange reistijd tot aan een sensor.

a) optimaal, t=1,7 uur

b) suboptimaal, t=2,2 uur



c) suboptimaal, t=2,8 uur

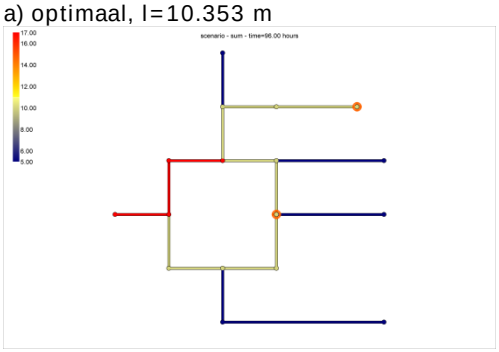
d) suboptimaal, t=5,8 uur

Figuur 23: Optimale en suboptimale sensorconfiguraties (oranje cirkels) voor de gemiddelde tijd tot de eerste detectie.

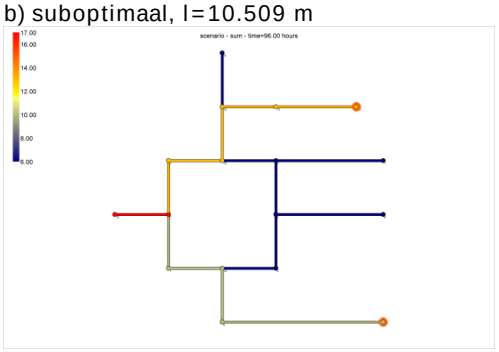
Criterium 4: Gemak van bronbepaling

Het gemak van bronbepaling drukken we uit in de gemiddelde omvang van het op basis van de sensormetingen te reconstrueren potentiële brongebied. Hierbij wordt voor alle besmettingsscenario's uitgegaan van de minimale omvang die in de periode na de besmetting wordt berekend. Figuur 24 toont een optimale en drie suboptimale configuraties voor een enkel besmettingsscenario (namelijk een 2 uur durende besmetting op t=72 uur). Voor besmettingsscenario's waarvoor geen detectie plaatsvindt, wordt de gehele netwerklengte omvang van het potentiële brongebied genomen (in feite een straflengte). Dit heeft tot gevolg dat een configuratie die op zichzelf een klein potentieel brongebied oplevert maar ook veel non-detecties, toch een grote gemiddelde omvang van het potentiële brongebied krijgt (voorbeeld Figuur 24d).

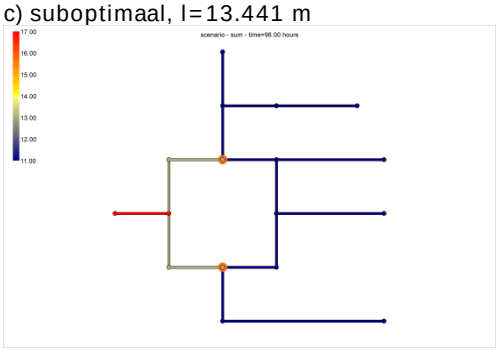
a) optimaal, l=10.353 m



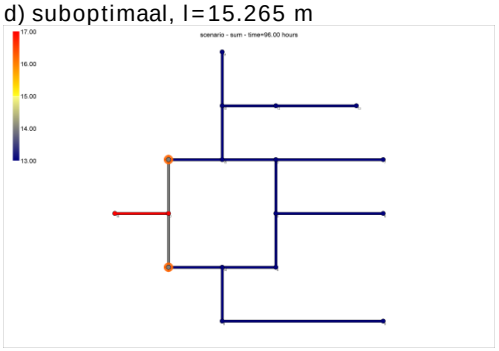
b) suboptimaal, l=10.509 m



c) suboptimaal, l=13.441 m



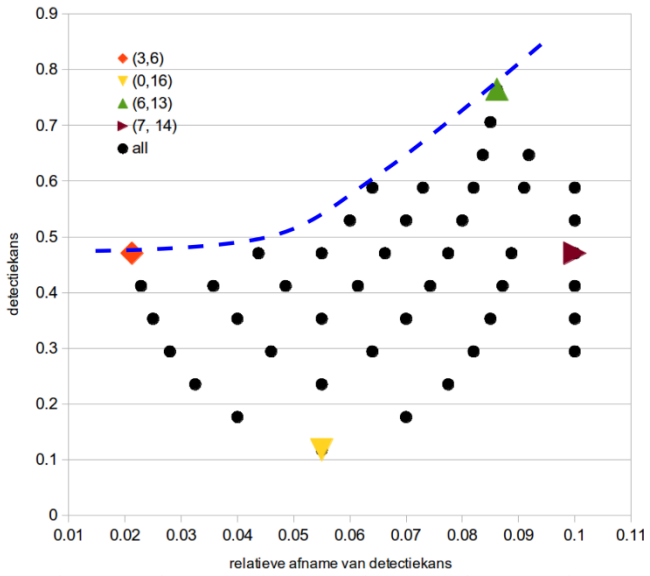
d) suboptimaal, l=15.265 m



Figuur 24: Optimale en suboptimale sensorconfiguraties (oranje cirkels) voor de gemiddelde minimale omvang van het potentiële brongebied.

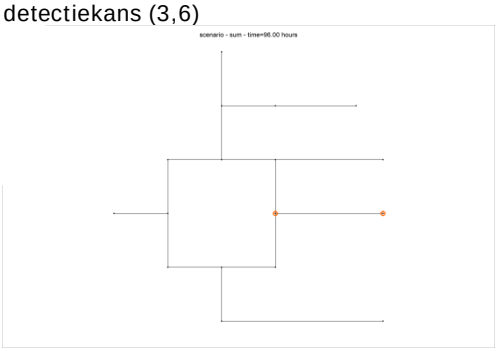
Criterium 5: Robuustheid m.b.t. detectiekans

De optimalisatiedoelstelling robuustheid met betrekking tot detectiekans is op zichzelf staand niet zo relevant. Een netwerk met 0 sensoren is immers bijzonder robuust tegen fout-negatieven, maar geeft geen relevante informatie. In Figuur 25 is de prestatie op deze robuustheid (bij faalkans 10%) voor alle mogelijk configuraties van 2 sensoren in het testnetwerk 1a uitgezet tegen de detectiekans zelf. Van de vier uithoeken van het diagram is een configuratie als illustratie weergegeven in Figuur 26. Omdat we alle mogelijke oplossingen hebben doorgerekend, vormt de curve van beste combinaties (hoge detectiekans, lage relatieve afname door vals-negatieve observaties) het Paretofront van optima.

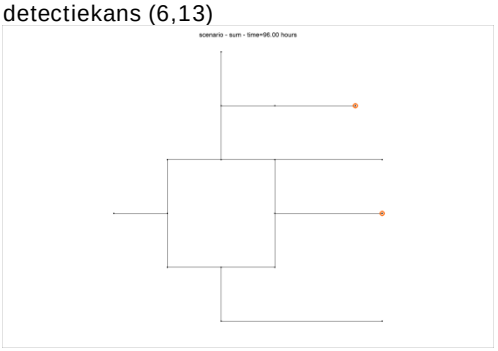


Figuur 25: Alle mogelijke oplossingen voor het plaatsen van 2 sensoren in het testnetwerk 1a geplot in de ruimte van relatieve afname van de detectiekans bij falen (kans op fout-negatief is 10%) tegen de detectiekans zelf. Vier oplossingen zijn specifiek aangeduid; hiervoor zijn de sensorconfiguraties weergegeven in Figuur 26.De blauwe onderbroken lijn geeft de ligging van het Paretofront van optimale combinaties aan.

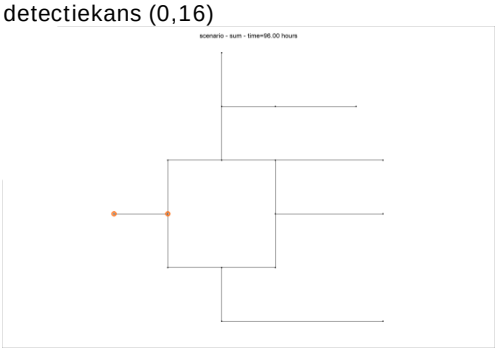
a) optimale robuustheid, suboptimale detectiekans (3,6)



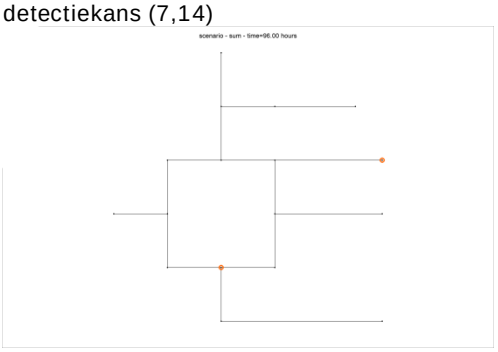
b) suboptimale robuustheid, optimale detectiekans (6,13)



c) suboptimale robuustheid, suboptimale detectiekans (0,16)



d) suboptimale robuustheid, suboptimale detectiekans (7,14)



Figuur 26: Optimale en suboptimale sensorconfiguraties (oranje cirkels) voor de robuustheid van het sensornetwerk voor fout-negatieve bepalingen m.b.t. het kunnen detecteren van besmettingen en detectiekansen.

Instelling van optimalisatie-algoritme

In de optimalisatie met het genetische algoritme wordt voor variatie in de populatie gebruik gemaakt van `crossover` op 2 punten gecombineerd met `network local mutation` (zie Tabel 2 op pagina 16). Voor het vaststellen van convergentie wordt de `change`-terminator gebruikt, waarbij wordt geëist dat 15 aansluitende generaties de verbetering van de beste prestatie minder dan een factor 10^{-5} is.

Tests en bepaling optimale waarden van overige parameters optimalisatie-algoritme

Tabel 16 geeft een overzicht van vergelijkende optimalisaties voor variërende combinaties van de relevante parameters die het genetische algoritme beheersen:

`population size`	aantal kandidaat-oplossingen dat in iedere generatie wordt beschouwd;
`network local mutation rate`	kans op een mutatie naar een andere knoop in de lokale netwerkomgeving
`network local mutation range`	bereik in de lokale netwerkomgeving van dergelijke mutaties (aantal knopen afstand)
`cross-over rate`	kans op een cross-over van twee kandidaat-oplossingen
`elite fraction`	fractie van een generatie die meegenomen wordt naar de volgende generatie (beste oplossingen)

Hierbij zijn de overige parameters zodanig gekozen dat het komen tot een oplossing niet-triviaal is. Steeds wordt 5 maal een optimalisatie uitgevoerd voor een combinatie van parameters en wordt het gemiddelde van de resultaten gerapporteerd. Deze resultaten zijn grafisch weergegeven in Figuur 27 voor verschillende combinaties van parameters. Groene

stippen geven combinaties van parameterwaarden aan die in alle onderzochte gevallen en herhalingen hebben geleid tot de globale optimale oplossing. Op basis hiervan zijn de volgende waarden gekozen voor alle overige optimalisaties die in dit rapport zijn beschreven:

population size	100
network local mutation rate	0,50
network local mutation range	8
cross-over rate	0,50
elite fraction	0,25

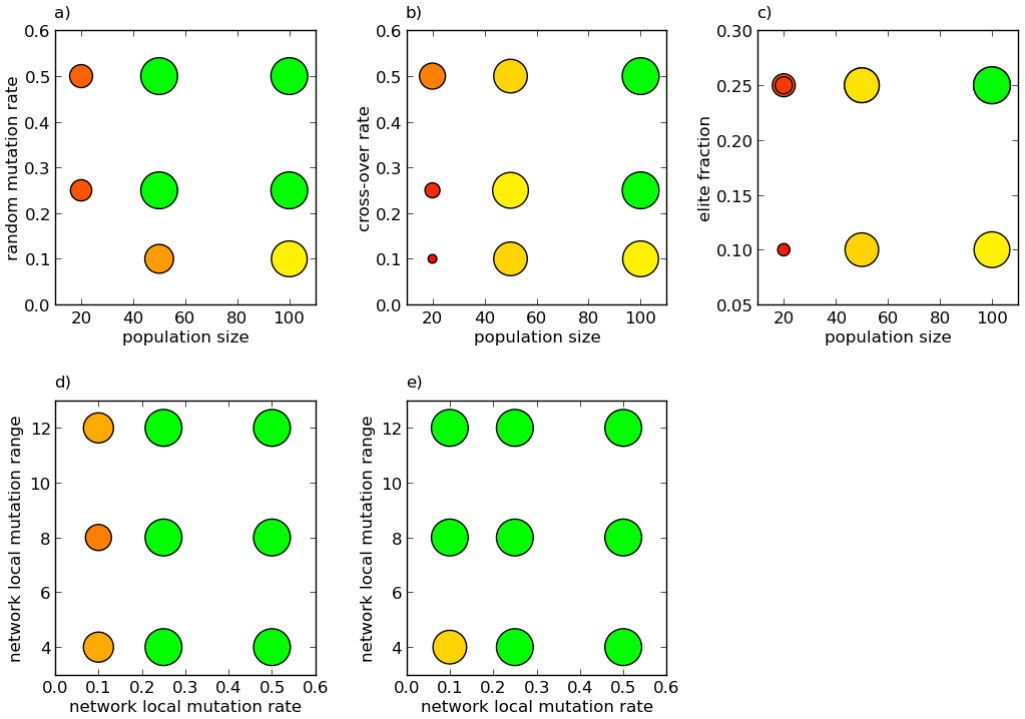
Tabel 16: Controle-experimenten van het genetisch algoritme. Voor verschillende combinaties van parameterwaarden voor het genetische algoritme wordt steeds 5 maal een optimalisatie uitgevoerd. De resulterende prestatie wordt vergeleken met het uit de brute-krachtberekening resulterende globale optimum (gele achtergrondkleur). De gekozen combinatie van parameterwaarden is gemarkeerd met een lichtblauwe achtergrondkleur.

	parameters van het genetisch algoritme					*resultaten*			
scenario	population size	network local mutation rate	network local mutation range	cross-over rate	elite fraction	*reken-tijd (sec)*	*aantal generaties*	*aantal evaluaties*	*prestatie: dekking (m)*
brute kracht			-			139.8	-	17030	97493,3
runs 1-5	20	0,1	4	0,1	0,1	4.47	32	640	89410,1
runs 6-10	20	0,1	8	0,1	0,1	3.81	25,8	516	91136,9
runs 11-15	20	0,1	12	0,1	0,1	4.94	36,2	724	91021,7
runs 16-20	20	0,25	4	0,1	0,1	3.91	27,6	552	85142,1
runs 21-25	20	0,25	8	0,1	0,1	4.46	31,4	628	93540,3
runs 26-30	20	0,25	12	0,1	0,1	4.71	34,6	692	92501,9
runs 31-35	20	0,5	4	0,1	0,1	3.43	24,4	488	92209,5
runs 36-40	20	0,5	8	0,1	0,1	4.88	34,8	696	94693,9
runs 41-45	20	0,5	12	0,1	0,1	4.12	28,2	564	94693,9
runs 46-50	50	0,1	4	0,1	0,1	7.84	30,2	1510	97493,3
runs 51-55	50	0,1	8	0,1	0,1	8.47	32,2	1610	94693,9
runs 56-60	50	0,1	12	0,1	0,1	5.75	22	1100	94974,7
runs 61-65	50	0,25	4	0,1	0,1	7.14	27	1350	97493,3
runs 66-70	50	0,25	8	0,1	0,1	6.80	25,8	1290	97493,3
runs 71-75	50	0,25	12	0,1	0,1	7.14	26,8	1340	97493,3
runs 76-80	50	0,5	4	0,1	0,1	6.21	23,2	1160	97493,3
runs 81-85	50	0,5	8	0,1	0,1	7.38	27,8	1390	97493,3
runs 86-90	50	0,5	12	0,1	0,1	6.68	25,4	1270	97493,3
runs 91-95	100	0,1	4	0,1	0,1	9.08	21,6	2160	94693,9
runs 96-100	100	0,1	8	0,1	0,1	10.65	25	2500	97493,3
runs 101-105	100	0,1	12	0,1	0,1	12.72	29,6	2960	97493,3
runs 106-110	100	0,25	4	0,1	0,1	9.88	22,4	2240	97493,3
runs 111-115	100	0,25	8	0,1	0,1	9.82	22,2	2220	97493,3
runs 116-120	100	0,25	12	0,1	0,1	10.45	23,6	2360	97493,3
runs 121-125	100	0,5	4	0,1	0,1	10.44	22,8	2280	97493,3
runs 126-130	100	0,5	8	0,1	0,1	11.07	23,4	2340	97493,3
runs 131-135	100	0,5	12	0,1	0,1	11.03	23,4	2340	97493,3
runs 136-140	20	0,1	4	0,1	0,25	3.26	22,8	456	84961,1
runs 141-145	20	0,1	8	0,1	0,25	4.25	29,8	596	84795,7

runs 146-150	20	0,1	12	0,1	0,25	4.66	31,4	628	89511,1
runs 151-155	20	0,25	4	0,1	0,25	4.49	31,6	632	91894,5
runs 156-160	20	0,25	8	0,1	0,25	4.18	28,4	568	89095,1
runs 161-165	20	0,25	12	0,1	0,25	3.72	26	520	91752,7
runs 166-170	20	0,5	4	0,1	0,25	3.95	26,6	532	94693,9
runs 171-175	20	0,5	8	0,1	0,25	3.63	25,4	508	91894,5
runs 176-180	20	0,5	12	0,1	0,25	4.31	28,8	576	97493,3
runs 181-185	50	0,1	4	0,1	0,25	6.09	22,6	1130	90061,3
runs 186-190	50	0,1	8	0,1	0,25	7.32	26,8	1340	97493,3
runs 191-195	50	0,1	12	0,1	0,25	8.54	30,8	1540	97493,3
runs 196-200	50	0,25	4	0,1	0,25	6.76	24,8	1240	97493,3
runs 201-205	50	0,25	8	0,1	0,25	6.15	22,4	1120	97493,3
runs 206-210	50	0,25	12	0,1	0,25	6.93	24,8	1240	97493,3
runs 211-215	50	0,5	4	0,1	0,25	5.87	21,8	1090	97493,3
runs 216-220	50	0,5	8	0,1	0,25	7.22	26,2	1310	97493,3
runs 221-225	50	0,5	12	0,1	0,25	7.44	26,4	1320	97493,3
runs 226-230	100	0,1	4	0,1	0,25	13.45	25,8	2580	97493,3
runs 231-235	100	0,1	8	0,1	0,25	12.74	24,4	2440	97493,3
runs 236-240	100	0,1	12	0,1	0,25	11.71	22,6	2260	97493,3
runs 241-245	100	0,25	4	0,1	0,25	12.46	23,4	2340	97493,3
runs 246-250	100	0,25	8	0,1	0,25	12.64	23,6	2360	97493,3
runs 251-255	100	0,25	12	0,1	0,25	12.32	23,2	2320	97493,3
runs 256-260	100	0,5	4	0,1	0,25	12.46	23,2	2320	97493,3
runs 261-265	100	0,5	8	0,1	0,25	12.32	22,6	2260	97493,3
runs 266-270	100	0,5	12	0,1	0,25	13.09	24	2400	97493,3
runs 271-275	20	0,1	4	0,25	0,1	5.61	38,2	764	95008,9
runs 276-280	20	0,1	8	0,25	0,1	5.94	39,6	792	90740,9
runs 281-285	20	0,1	12	0,25	0,1	4.85	32,6	652	92144,9
runs 286-290	20	0,25	4	0,25	0,1	3.89	25,8	516	94693,9
runs 291-295	20	0,25	8	0,25	0,1	4.06	26,8	536	93540,3
runs 296-300	20	0,25	12	0,25	0,1	4.66	30,6	612	97493,3
runs 301-305	20	0,5	4	0,25	0,1	3.84	25	500	87261,9
runs 306-310	20	0,5	8	0,25	0,1	5.18	33,8	676	97493,3
runs 311-315	20	0,5	12	0,25	0,1	6.13	39,8	796	94693,9
runs 316-320	50	0,1	4	0,25	0,1	8.37	28,4	1420	97493,3
runs 321-325	50	0,1	8	0,25	0,1	6.46	21,6	1080	94693,9
runs 326-330	50	0,1	12	0,25	0,1	8.93	30,2	1510	97493,3
runs 331-335	50	0,25	4	0,25	0,1	6.22	20,2	1010	97493,3
runs 336-340	50	0,25	8	0,25	0,1	7.64	24,6	1230	97493,3
runs 341-345	50	0,25	12	0,25	0,1	8.36	27,2	1360	97493,3
runs 346-350	50	0,5	4	0,25	0,1	9.03	28,8	1440	97493,3
runs 351-355	50	0,5	8	0,25	0,1	8.49	26,4	1320	97493,3
runs 356-360	50	0,5	12	0,25	0,1	8.13	25,2	1260	97493,3
runs 361-365	100	0,1	4	0,25	0,1	13.17	23	2300	97493,3
runs 366-370	100	0,1	8	0,25	0,1	13.31	23,2	2320	97493,3
runs 371-375	100	0,1	12	0,25	0,1	14.10	24,4	2440	97493,3
runs 376-380	100	0,25	4	0,25	0,1	13.70	23,6	2360	97493,3
runs 381-385	100	0,25	8	0,25	0,1	13.04	22	2200	97493,3
runs 386-390	100	0,25	12	0,25	0,1	13.65	23	2300	97493,3
runs 391-395	100	0,5	4	0,25	0,1	12.65	21,6	2160	97493,3

runs 396-400	100	0,5	8	0,25	0,1	14.67	24,8	2480	97493,3
runs 401-405	100	0,5	12	0,25	0,1	13.83	23,4	2340	97493,3
runs 406-410	20	0,1	4	0,25	0,25	4.23	24,8	496	91894,5
runs 411-415	20	0,1	8	0,25	0,25	7.09	39	780	94693,9
runs 416-420	20	0,1	12	0,25	0,25	5.64	32,8	656	96454,9
runs 421-425	20	0,25	4	0,25	0,25	5.35	31	620	94693,9
runs 426-430	20	0,25	8	0,25	0,25	4.92	27	540	94693,9
runs 431-435	20	0,25	12	0,25	0,25	4.61	25,4	508	97493,3
runs 436-440	20	0,5	4	0,25	0,25	4.52	25,2	504	92209,5
runs 441-445	20	0,5	8	0,25	0,25	5.34	29,2	584	93540,3
runs 446-450	20	0,5	12	0,25	0,25	4.97	27,4	548	94693,9
runs 451-455	50	0,1	4	0,25	0,25	9.84	28,6	1430	97493,3
runs 456-460	50	0,1	8	0,25	0,25	10.23	28,4	1420	97493,3
runs 461-465	50	0,1	12	0,25	0,25	10.18	29,2	1460	97493,3
runs 466-470	50	0,25	4	0,25	0,25	9.29	25,4	1270	97493,3
runs 471-475	50	0,25	8	0,25	0,25	8.32	23	1150	97493,3
runs 476-480	50	0,25	12	0,25	0,25	8.33	22,6	1130	97493,3
runs 481-485	50	0,5	4	0,25	0,25	9.81	25,8	1290	97493,3
runs 486-490	50	0,5	8	0,25	0,25	9.04	24,2	1210	97493,3
runs 491-495	50	0,5	12	0,25	0,25	9.05	23,8	1190	97493,3
runs 496-500	100	0,1	4	0,25	0,25	14.98	22,4	2240	97493,3
runs 501-505	100	0,1	8	0,25	0,25	15.17	23,2	2320	97493,3
runs 506-510	100	0,1	12	0,25	0,25	14.72	22,6	2260	97493,3
runs 511-515	100	0,25	4	0,25	0,25	13.89	21	2100	97493,3
runs 516-520	100	0,25	8	0,25	0,25	13.74	21	2100	97493,3
runs 521-525	100	0,25	12	0,25	0,25	14.36	21,4	2140	97493,3
runs 526-530	100	0,5	4	0,25	0,25	14.24	21,2	2120	97493,3
runs 531-535	100	0,5	8	0,25	0,25	15.30	22,2	2220	97493,3
runs 536-540	100	0,5	12	0,25	0,25	17.90	26	2600	97493,3
runs 541-545	20	0,1	4	0,5	0,1	5.90	27,6	552	90376,3
runs 546-550	20	0,1	8	0,5	0,1	6.79	31,8	636	94693,9
runs 551-555	20	0,1	12	0,5	0,1	6.74	33,2	664	96013,1
runs 556-560	20	0,25	4	0,5	0,1	7.02	31,8	636	97493,3
runs 561-565	20	0,25	8	0,5	0,1	6.55	28,8	576	97493,3
runs 566-570	20	0,25	12	0,5	0,1	6.29	29,4	588	97493,3
runs 571-575	20	0,5	4	0,5	0,1	6.33	28,8	576	97493,3
runs 576-580	20	0,5	8	0,5	0,1	6.55	29,8	596	97493,3
runs 581-585	20	0,5	12	0,5	0,1	5.59	26,6	532	97493,3
runs 586-590	50	0,1	4	0,5	0,1	11.03	27,8	1390	97493,3
runs 591-595	50	0,1	8	0,5	0,1	8.01	20,6	1030	97493,3
runs 596-600	50	0,1	12	0,5	0,1	9.75	24,4	1220	96454,9
runs 601-605	50	0,25	4	0,5	0,1	9.17	23,2	1160	97493,3
runs 606-610	50	0,25	8	0,5	0,1	10.79	27	1350	97493,3
runs 611-615	50	0,25	12	0,5	0,1	11.06	27,4	1370	97493,3
runs 616-620	50	0,5	4	0,5	0,1	9.10	22	1100	97493,3
runs 621-625	50	0,5	8	0,5	0,1	10.25	24,8	1240	97493,3
runs 626-630	50	0,5	12	0,5	0,1	10.73	25,8	1290	97493,3
runs 631-635	100	0,1	4	0,5	0,1	16.49	23,2	2320	97493,3
runs 636-640	100	0,1	8	0,5	0,1	14.89	20,6	2060	97493,3
runs 641-645	100	0,1	12	0,5	0,1	16.44	22	2200	97493,3

runs 646-650	100	0,25	4	0,5	0,1	14.91	19,8	1980	97493,3
runs 651-655	100	0,25	8	0,5	0,1	15.94	20,6	2060	97493,3
runs 656-660	100	0,25	12	0,5	0,1	15.61	20,6	2060	97493,3
runs 661-665	100	0,5	4	0,5	0,1	16.79	22,6	2260	97493,3
runs 666-670	100	0,5	8	0,5	0,1	16.71	22,4	2240	97493,3
runs 671-675	100	0,5	12	0,5	0,1	17.30	22,6	2260	97493,3
runs 676-680	20	0,1	4	0,5	0,25	6.97	31	620	94693,9
runs 681-685	20	0,1	8	0,5	0,25	6.12	28,4	568	92860,7
runs 686-690	20	0,1	12	0,5	0,25	5.84	25,2	504	89322,9
runs 691-695	20	0,25	4	0,5	0,25	6.45	30	600	97493,3
runs 696-700	20	0,25	8	0,5	0,25	6.07	25,8	516	94693,9
runs 701-705	20	0,25	12	0,5	0,25	6.24	28,4	568	96454,9
runs 706-710	20	0,5	4	0,5	0,25	5.82	25,4	508	94693,9
runs 711-715	20	0,5	8	0,5	0,25	6.15	27,2	544	97493,3
runs 716-720	20	0,5	12	0,5	0,25	7.16	30	600	97493,3
runs 721-725	50	0,1	4	0,5	0,25	9.97	22,6	1130	95008,9
runs 726-730	50	0,1	8	0,5	0,25	11.90	27	1350	96424,5
runs 731-735	50	0,1	12	0,5	0,25	13.83	31,2	1560	97493,3
runs 736-740	50	0,25	4	0,5	0,25	11.28	25,6	1280	97493,3
runs 741-745	50	0,25	8	0,5	0,25	9.43	20,8	1040	97493,3
runs 746-750	50	0,25	12	0,5	0,25	10.74	23,8	1190	97493,3
runs 751-755	50	0,5	4	0,5	0,25	10.23	23,2	1160	97493,3
runs 756-760	50	0,5	8	0,5	0,25	11.34	25,2	1260	97493,3
runs 761-765	50	0,5	12	0,5	0,25	10.69	23,6	1180	97493,3
runs 766-770	100	0,1	4	0,5	0,25	17.03	21,2	2120	97493,3
runs 771-775	100	0,1	8	0,5	0,25	17.51	21,4	2140	97493,3
runs 776-780	100	0,1	12	0,5	0,25	19.84	24	2400	97493,3
runs 781-785	100	0,25	4	0,5	0,25	17.28	21	2100	97493,3
runs 786-790	100	0,25	8	0,5	0,25	17.95	21,6	2160	97493,3
runs 791-795	100	0,25	12	0,5	0,25	18.27	22,2	2220	97493,3
runs 796-800	100	0,5	4	0,5	0,25	17.65	21,2	2120	97493,3
runs 801-805	100	0,5	8	0,5	0,25	19.24	23	2300	97493,3
runs 806-810	100	0,5	12	0,5	0,25	19.90	23,6	2360	97493,3



Figuur 27: Prestatie van de optimalisatie met het genetische algoritme in vergelijking met de brute-krachtoptimalisatie als functie van verschillende combinaties van GA-parameters. Groene stippen geven volledige overeenkomst weer; in de andere gevallen geeft de kleur (rood=0 -> geel=1) en het oppervlak van de stip de fractie aan van de oplossingen die overeenkomen met de brute-krachtoplossing. a) Populatiegrootte tegen willekeurige mutatiefrequentie; b) populatiegrootte tegen crossoverfrequentie; c) populatiegrootte tegen elitefractie; d) mutatiefrequentie en –bereid voor network local mutation met populatiegrootte 50; e) mutatiefrequentie en –bereid voor network local mutation met populatiegrootte 100.

Bijlage III: Beschrijving en benchmark van ontwikkelde tool voor optimale sensorlocatie-keuze

Beschrijving van voorbereidende en ondersteunende programma's

Tabel 17 geeft een overzicht van relevante ondersteunende programma's. Beide programma's zijn geschreven in Python. Zie ook Van Thienen en Vries (2013).

Tabel 17: Overzicht van voorbereidende en ondersteunende tools.

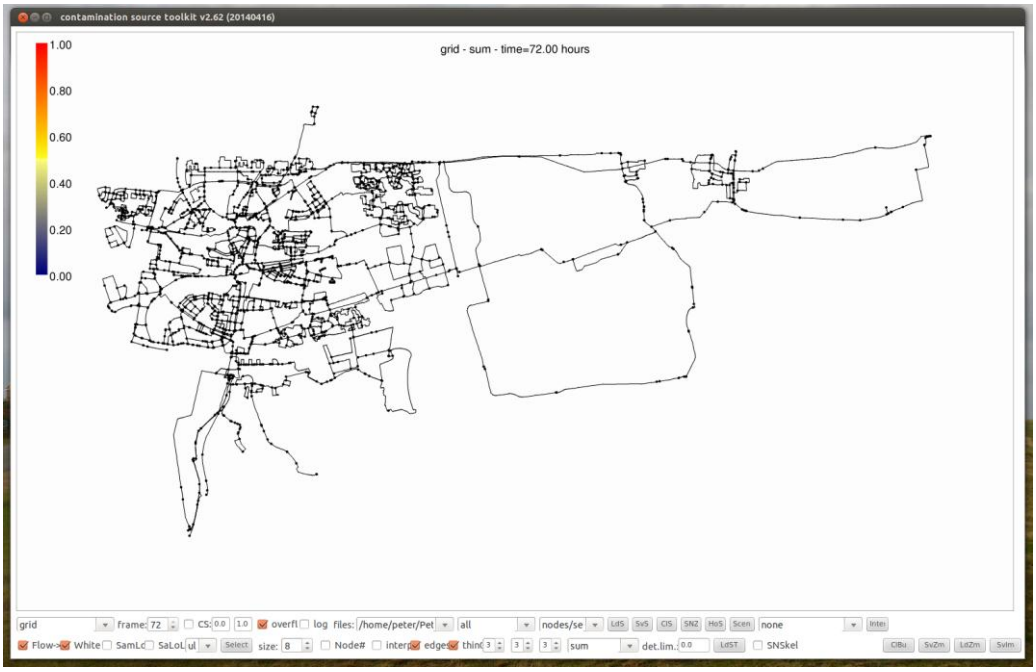
naam	doel	type
`EPANETinp2msx`	Maakt op basis van een EPANET-invoerbestand een reeks invoerbestanden voor EPANET-MSX t.b.v. backtraceberekeningen	command line
`joinmsxresults`	Voegt de resultaten van een reeks berekeningen met EPANET-MSX samen tot een enkel bestand met interne compressie. Hiermee wordt de omvang van de data met een factor 25 teruggebracht en het verwerken van de resultaten aanzienlijk versneld.	command line

Beschrijving van de Contamination Source Toolkit

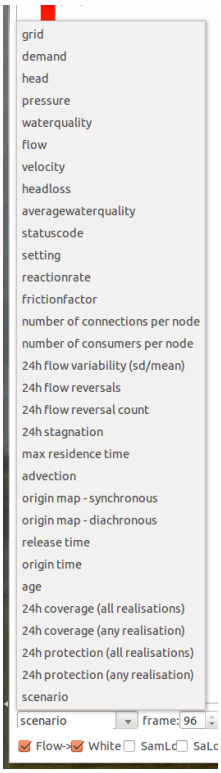
De *Contamination Source Toolkit* (CST) is een programma voor de analyse van de herkomst van besmettingen, het doorrekenen van besmettingsscenario's en de bepaling van optimale locaties voor waterkwaliteitssensoren op basis van diverse doelstellingen. Het is gebaseerd op in een eerder project geschreven software (Van Thienen en Vries, 2013) en uitgebreid in dit project. Het programma is geschreven in Python, waarmee het in principe op diverse platformen te gebruiken is (hoewel uitsluitend getest en toegepast op Linux). Tabel 18 geeft een overzicht van de functionaliteit en mogelijkheden van het programma. Dit is niet bedoeld om compleet te zijn, maar slechts om een indruk te geven.

Tabel 18: Overzicht van functionaliteiten van CST.

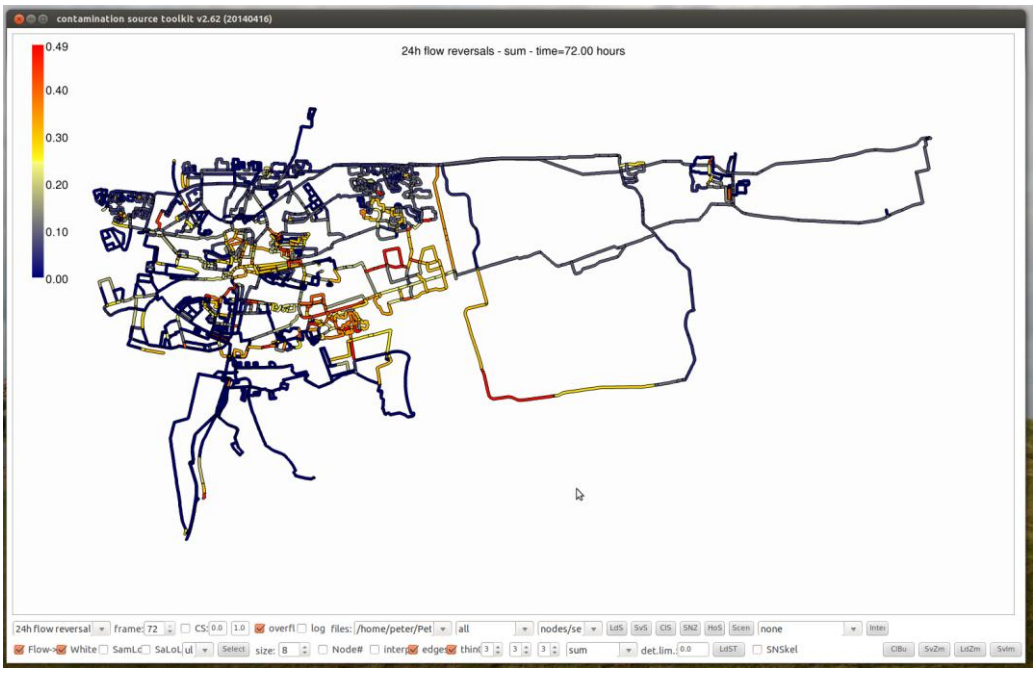
Functionaliteit	*Beschrijving*	*Figuur*
Basisscherm	Visualisaties, interacties, toegang tot complexere analyses, invoer, uitvoer	Figuur 28
Directe visualisaties	Menu waarin directe visualisatie van allerlei hydraulische en waterkwaliteitsinformatie kan worden geselecteerd	Figuur 29
Voorbeeld	Visualisatie van fractie van de tijd dat water in tegengestelde richting stroomt.	Figuur 30
Scenario's en optimalisaties	Instellingenvenster voor het uitvoeren van scenarioberekeningen en optimalisaties	Figuur 31
Operatie	Type scenario of optimalisatie dat moet worden uitgevoerd.	Figuur 32a
Doelstelling	Doelstelling van optimalisatie	Figuur 32b
Algoritme	Te gebruiken optimalisatie-algoritme	Figuur 32c
Paretovenster	Visualisatie van de prestatie van een populatie van sensorconfiguraties voor twee doelstellingen	Figuur 33



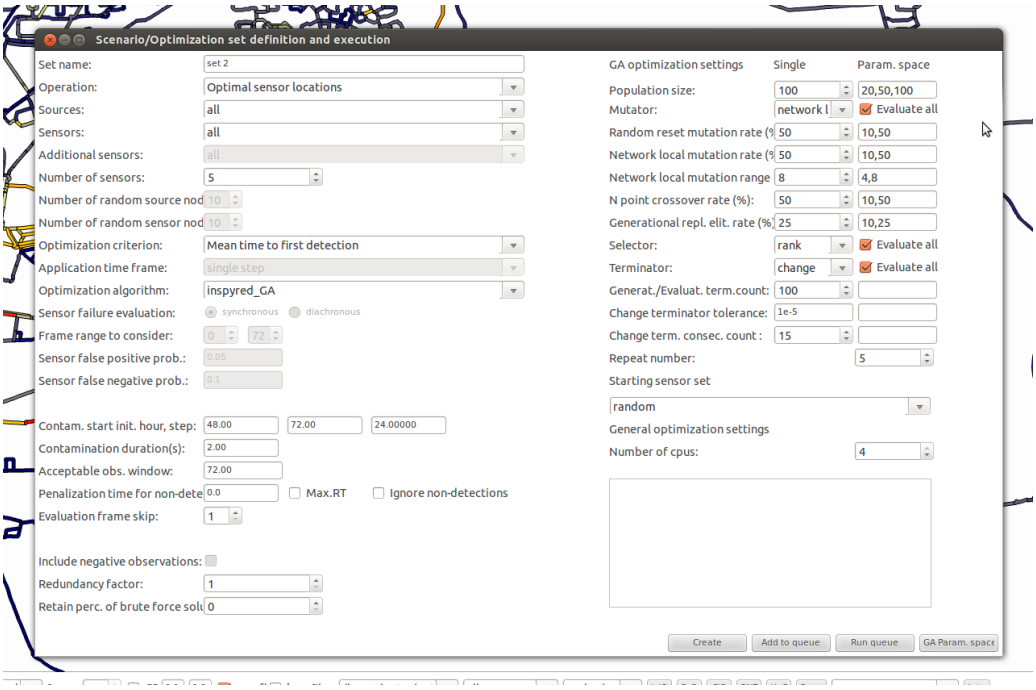
Figuur 28



Figuur 29

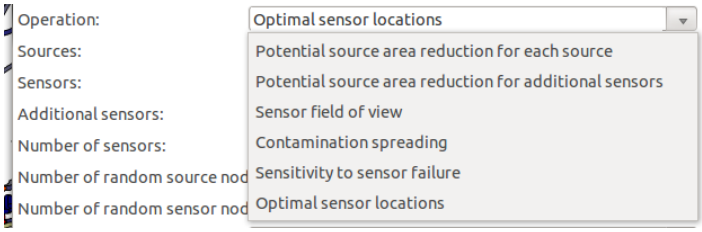


Figuur 30

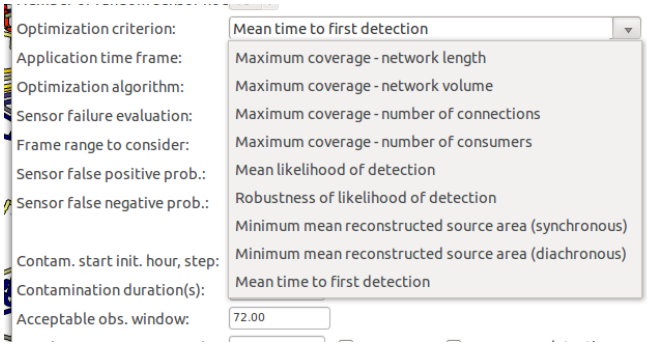


Figuur 31

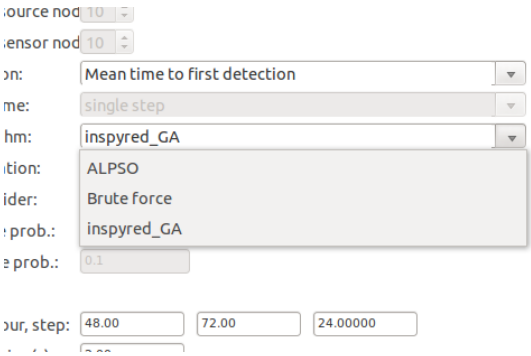
a)



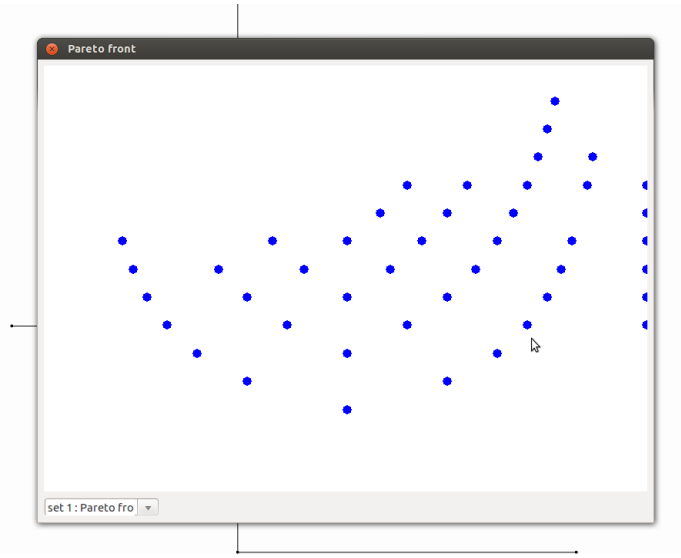
b)



c)



Figuur 32



Figuur 33

Stappenplan voor evaluatie en optimalisatie met cst en bijbehorende tools

Voor het uitvoeren van de in dit proces beschreven analyses dient een aantal stappen te worden doorlopen:

1. Voorbereiden (inclusief eventueel verdunnen) en testen van EPANET2-netwerkmodel (`.inp`-bestand).
2. Genereren van EPANET-MSX-invoerbestand op basis van het `.inp`-bestand met behulp van `EPANETinp2msx`.
3. Uitvoeren van hydraulische simulaties (EPANET2) en simulaties van stoftransport (EPANET-MSX).
4. Samenvoegen en compressie van resultaten transportberekeningen met `joinmsxresults`.
5. Inlezen, selectie potentiële bron- en sensorlocaties en analyse van resultaten met `cst`.

Benchmark

Als benchmark zijn prestatieberekeningen van enkele geoptimaliseerde sensornetwerken uit de Battle of the Water Sensor Networks (Ostfeld et al., 2008) opnieuw uitgevoerd met cst. De resultaten zijn weergegeven in Tabel 19. In twee van de drie gevallen komt de door `cst` bepaalde pakkans zeer goed overeen met de gerapporteerde waarde, in het derde geval is de waarde iets lager. Voor de detectietijd blijkt dat `cst` steeds met een 10-20% lagere waarde komt dan de gerapporteerde. Het is moeilijk te bepalen waar de verschillen aan toe te schrijven zijn. Reeds eerder is gebleken dat doorrekening van eenzelfde netwerk met verschillende hydraulische codes tot andere resultaten kan leiden (Van Summeren, 2014). De in Bijlage II gerapporteerde testresultaten geven voldoende vertrouwen in de uitvoer van `cst`. Wel duiden deze bevindingen erop dat men voorzichtig moet zijn met het vergelijken van prestaties van sensornetwerken die met verschillende methoden of programma's zijn berekend.

Tabel 19: Vergelijking van prestatieberekeningen voor geselecteerde sensornetwerken uit Ostfeld et al. (2008) zoals daar gerapporteerd en zoals bepaald met cst.

		Gerapporteerd in Ostfeld et al. (2008)		Bepaling met code van dit rapport	
oplossing	sensorlocaties	detectie-tijd (min.)	pakkans	detectie-tijd (min.)	pakkans
Berry et al.	17, 21, 68, 79, 122	542	0,609	439	0,598
Ostfeld en Salomons	117, 71, 98, 68, 82	461	0,622	420	0,621
Wu en Walski	45, 68, 83, 100, 118	704	0,787	604	0,718